Profiling of ABC transporters

Wolters, Justina Clarinda

Profiling of ABC transporters

Justina Clarinda Wolters

Cover: Karin Wolters

Printed by: PrintPartners Ipskamp, Enschede, The Netherlands

ISBN: 978-90-376-4571-0

This Ph.D. study was carried out in the Biochemistry Department of the Groningen Biomolecular Sciences and Biotechnology Institute (GBB) of the University of Groningen and was financially supported by the Netherlands Proteomics Centre (NPC).

RIJKSUNIVERSITEIT GRONINGEN

Profiling of ABC transporters

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
vrijdag 15 oktober 2010
om 14:45 uur

door

Justina Clarinda Wolters
geboren op 2 december 1981
te Noordoostpolder

Promotor:	Prof. Dr. B. Poolman
Beoordelingscommissie:	Prof. Dr. Ir. A.J. Minnaard Prof. Dr. J. Kok Prof. Dr. E.J. Boekema

Contents

Chapter 1 An introduction to adenine nucleotide-binding proteins

Adenosine triphosphate

The nucleotide adenosine triphosphate (ATP) is involved in many aspects of cellular activity. It is a main source of metabolic energy in living cells, fuelling transport (via transport ATPases) or movement (via molecular motors), phosphorylation of enzymes and small metabolites via kinases; in general ATP is used to drive many endergonic reactions. Moreover, it is a substrate and/or cofactor involved in DNA/RNA modifying processes like transcription, DNA replication, recombination and restriction. Besides its function as energy storage unit or substrate for the biosynthesis of DNA/RNA, ATP acts as a signaling molecule for proteins belonging to the class of purinergic receptors. These different classes of adenine nucleotide-binding proteins will be introduced shortly; a more complete overview is presented in Chapter 2. Within these different classes, the subclass of ATP binding cassette (ABC) transport ATPases has been the focus of this thesis and these proteins will be described in most detail.

Functional grouping of ATPases

Although numerous nucleotide-utilizing proteins can be categorized on the basis of four basic folds (as will be discussed below), a number of proteins do not belong to these four fold-classes. Proteins like the aminoacyl-tRNA synthetases, (GTP hydrolyzing) tubulin, proteins that use nucleotides as building blocks (like polymerases) or signaling molecules (purinergic receptors) do have a distinctly different structure [1]. For this reason, a functional grouping was done in the following categories: molecular motors, kinases, RNA and DNA modifying enzymes, signaling and transport ATPases.

Molecular motors

Molecular motors use ATP binding and hydrolysis to move along cytoskeletal elements and fall in three categories: myosin, kinesin and dynein. Myosin is involved in the movement of muscle and moves over actin-based polymers, while kinesin is involved in the movement of membrane organelles along microtubules. Both motors have a different function and distinct mechanochemical cycles, but a similar core structure where ATP hydrolysis is translated to mechanical movement via the conserved γ-phosphate sensor [2]. Dynein uses microtubule for movement like the kinesins, but in the opposite direction (towards the minus-end of the microtubule) and functions for the transport of cargo, but also for the beating movement of flagella or cilia [3]. Dynein is significantly different from myosin and kinesin and the motor domain of this protein is composed of six ATPases that cluster with the AAA-family (ATPases associated with a variety of cellular activities) [4].

Kinases

Kinases form a large and diverse group of adenine nucleotide-binding proteins and use ATP to phosphorylate a wide variety of substrates ranging from small molecules to proteins and names are derived from these different acceptors. The classes of protein kinases, for example, are named after the residue that is phosphorylated within the protein like serine/threonine and tyrosine kinases. Sugar kinases are named after the specific sugar that is phosphorylated, for example, fructokinase and hexokinase. This phosphorylation can serve a role in catalysis or signal transduction.

RNA and DNA modifying enzymes

Nucleotides (NTP) are required as substrates in the biosynthesis of RNA (NTP) and DNA (deoxynucleotides (dNTP)) by polymerases. Ribunucleotide reductases convert the nucleotides (in their diphosphate or triphosphate form) to their corresponding deoxyribonucleotides. A subset of ribonucleotide reductases contain an active-site, where ATP binding acts as an activator of the reductase activity, while dATP binding on this site acts as an inhibitor [5]. The different functions of ATP as building blocks, as cofactor or as source of metabolic is seen in DNA replication, repair and transcription, involving DNA and RNA polymerases, helicases, topoisomerases, recombinases, transcription factors and tRNA synthetases. RNA polymerase is involved in transcription of RNA from a DNA template strand. Transcription initiation of mRNA is started by a complex of the RNA polymerase type II and a number of general transcription factors (TF). ATP hydrolysis is required for the initiation and promoter escape, and this requirement is linked to one of the transcription factors (TFIIH) that contains helicase and kinase activity [6]. After subsequent elongation, transcription termination (in bacteria) can be done by the ATP-utilizing transcription factor Rho (Rho-dependent pathway), although Rho-independent pathways exist as well. Rho is an ATPase closely related to the F-type ATPases and uses ATP hydrolysis for transcript release [7].

DNA polymerases are involved in replication of DNA using a template strand. The replicative DNA polymerases are loaded onto the DNA via protein complexes called clamp loaders, and ATP hydrolysis is needed to form the complex around the DNA [8,9]. The subunits of the clamp loader are members of the AAA-ATPase family (ATPases associated with a variety of cellular activities). Several other members of the AAA-ATPase family are involved in the initiation of DNA replication, e.g. proteins like the replication initiator and the helicase loader. The replication initiator binds to the origin of replication, using ATP binding for limited unfolding of the DNA strand. The helicase loader uses the ATP binding for loading of helicases [8], whereas helicases use ATP hydrolysis as energy source for the splitting of two complementary strands [10]. The coiled or super-coiled topology of DNA strands needs to be reduced to allow movement of the helicases over the DNA strands, which is done by topoisomerases. Depending on the topology of the DNA, DNA topoisomerases can relax supercoiling, unlink daughter replicons or unknot DNA strands [11], which are important processes for DNA replication

and transcription. The RecA ATPase protein (in bacteria) is also involved in a number of DNA-related processes, like DNA repair, induction of a SOS response upon extensive DNA damage and SOS mutagenesis activity, and its main function is to homologously recombine DNA strands [12].

tRNA synthetases use ATP as source of metabolic energy to charge transfer RNAs (tRNA). The ATP hydrolysis is used to form the amino-adenylate complexes from ATP and an amino acid, and this complex subsequently reacts with the tRNA to form the aminoacyl-tRNA [13]. The aminoacyl-tRNAs are recognized by the elongation factors to recruit the tRNAs to the ribosome for translation. The bacterial elongation factor (EF-Tu) is unable to use ATP, instead it uses GTP to bind aminoacyl-tRNAs to the ribosome [14]. Yeast contains an additional elongation factor (besides the two homologs EF-1α and Ef-2 that show conservation with the two bacterial homologs EF-Tu and EF-G), namely EF-3. The ATPase EF-3 stimulates the binding of EF-1α, and both ATP and EF-3 are required for translation [15].

Signaling

ATP is an extracellular signaling molecule that is detected by purinergic receptors, but ATP can also serve as substrate for the production of other (intracellular) signaling molecules, like cyclic AMP (cAMP) via adenylate cyclase (also named adenylyl cyclase). Purinergic receptors can be classified by their sensitivity into two categories: P2-purinergic receptors respond to ATP>ADP>AMP>adenosine, while P1-purinergic receptors show a reverse sensitivity being mostly activated by adenosine and least activated by ATP [16]. The class of P2-purinergic receptors can be subdivided based on the way the signal is transduced: the class of P2Y-purinergic receptors signal through G-protein coupled receptors (GPCRs) and the P2X-purinergic receptors signal through ligand-gated ion channels [16].

Transport ATPases

The P-, V- and F-ATPases together with the ABC transporters belong to the cluster of transport ATPases [17]. P-ATPases are involved in the transport of various cations and metals. The P-ATPases, have a covalent phosphorylated intermediate in their enzymatic cycle (hence the name P-ATPases), which sets them apart from the other types of transport ATPases [18]. The activity of the V- and F-ATPases is coupled to the formation of an electrochemical ion (proton or sodium ion) gradient. V- ATPases use ATP hydrolysis to generate an ion gradient, while F-ATPases can also work in the reversion direction and synthesize ATP (ATP synthases) at the expense of an electrochemical ion gradient [19]. The class of ABC transporters is the largest (super)family within the transport ATPases. The members are found in all kingdoms of life and play an important role in a wide range of physiological functions, mostly but not exclusively involving translocation of compounds across the membrane, both as importers and exporters [20,21]. The class of ABC transporters will be discussed more extensively below.

Structural grouping of ATPases

Among the ATP-utilizing enzymes, a number of proteins can be clustered based on their fold, namely the fold of the ASCE proteins (part of the P-loop NTPases), the fold of GHKL proteins, the actin-fold proteins and the fold of chaperonins [1].

ASCE proteins

ASCE (additional strand conserved E) proteins are part of the superfamily of P-loop NTPases and these proteins share a characteristic fold in the core domain of (at least) five parallel β-sheets, which are sandwiched between two layers of α-helices (Figure 1B). The P-loop NTPases are clustered together by the presence of the P-loop (phosphate-binding loop), which is used in the binding of the nucleotides. The P-loop is also known as the Walker A motif and has a conserved sequence GxxxxGK[ST] (x can be any amino acid) [22,23]. The second motif present in this superfamily is the Walker B motif with the general sequence ϕϕϕϕD (ϕ can be any hydrophobic amino acid). The ASCE class within the P-loop NTPases contains an additional β-strand between the P-loop and the Walker B motif, and the presence of a conserved acidic residue (usually glutamate), which defines the preference for ATP. These two characteristics set the ASCE proteins apart from the other class of P-loop NTPases, namely the KG proteins (the nucleotide kinases and GTPases). The ASCE proteins can be subdivided in the following subclasses: the AAA-, ABC-, RecA-like ATPases, SFI/II helicases, FtsK, PilT, KAP and STAND categories (reviews [24,25]), which are characterized by additional secondary structures or domains and the presence of conserved residues at specific positions). The relationships between these classes and subclasses are visualized in Figure 1A.

The ABC, RecA-like, SFI/II helicases, PilT and FtsK are more related to each other than to other proteins due to the presence of a C-terminal hairpin next to the core domain, which is formed because these proteins contain an additional antiparallel β-strand at the C-terminus. ABC (ATP Binding Cassette) transporters are involved in the transport of a large variety of substrates over the membrane and a small subgroup (sometimes named ABC proteins) is involved in DNA repair or translation elongation [20]. The RecA-like (Recombinase A-like) ATPases are involved in several DNA modifying processes, with a main function as DNA recombinase [12]; the F-ATPases also belong to this group. SF1/SF2 (superfamily I/II) helicases are a cluster formed by the majority of the helicases (a small number of helicases group in superfamily III and belong to the AAA-ATPases) and are important for RNA splicing, DNA repair and recombination, transcription and translation [10]. The superfamiliy of FtsK is a class of hexameric ring-forming ATPases, named after the cell division protein FstK, and also contains proteins that form pores for the extrusion of DNA and peptides. This class of proteins clusters together with the PilT (pilus-dependent surface motility retraction motor [26]) class of proteins based on PSI-BLAST analyses, but the structural or sequential basis for this clustering is yet unkown [25]. The RecA-like, SF1/2 helicases, PilT and FtsK ATPases

cluster together relative to the ABC transporters by the presence an additional β-strand after the second β-strand of the core domain [25].

The second cluster is formed by AAA-, KAP- and STAND-ATPases, which are grouped based on the presence of an additional C-terminal helical bundle. AAA-ATPases (ATPases associated with a variety of cellular activities) are involved in processes like DNA replication, protein folding and degradation, assembly or disassembly of complexes of polypeptides or nucleic acids as well as movement (dynein) [27]. The AAA-ATPases are defined by the absence of β-strands apart from those in the core domain and the presence of a small helical bundle fused to the C- terminus of the core domain [28]. The KAP-family of proteins regulate NTP-dependent (dis)assembly of protein complexes and are named after its members Kidins 220 (kinase D-interacting substance of 220 kDa) /ARMS (ankyrin repeat-rich membrane spanning protein) and PifA (a F-plasmid protein involved in phage T7 exclusion). The family contains an insert in the P-loop NTPase domain of two or four transmembrane helices [24]. STAND-ATPases (signal transduction ATPases with numerous domains) are involved in processes like apoptosis, plant pathogenesis and bacterial translational regulation and this class is related to the AAA-ATPases. The domains present in addition to the NTPase units usually comprise interaction domains (for DNA or protein binding), superstructure-forming repeats and enzymatic domains involved in signal transduction [29,30]. The specific motifs of each of these classes are not further discussed within this introduction, but they can be found in the references mentioned heretofore for the distinctive classes of P-loop NTPases.

GHKL fold

The GHKL fold named by the ATPases with this specific nucleotide-binding fold, namely the DNA gyrase, heat shock protein 90, histidine kinase and the DNA repair protein MutL. The different members share a highly conserved ATP-binding domain that is linked to a non-conserved dimerization domain. The ATP-binding domain is formed by a layer of four antiparallel β-sheets with a second layer formed by three α-helices, which is different from P-loop NTPase fold where α-helices surround parallel β-sheets on both sides (Figure 1B) [31].

Actin fold

Proteins containing the actin-type fold use the energy derived from ATP to convert the conformational changes into directed action, for example in the polymerization of monomeric actin subunits into filaments by actin, the function of chaperones in protein folding, remodeling of protein complexes, intracellular transport, and the phosphorylation of various acceptors by kinases [1,32]. The class contains actin, heat shock protein 70 (Hsc70), sugar and glycerol kinases; a number of cell cycle proteins and two phosphatases are also predicted to fall in this family [32,33]. The basic actin fold is formed by a structure of two domains, each subdivided in a large and small (less conserved) subdomain. ATP is bound at the interface of the two large domains, which

contain five mixed β-strands surrounded by three helices (Figure 1B); the ATP binding motifs are significantly different from the motifs in the P-loop NTPases [1,32].

Chaperonins

Chaperonins, like GroEL, are formed by oligomeric structures in which ATP hydrolysis leads to large conformational changes and promotes unfolding of protein substrates. A GroEL protomer is composed of three domains: an equatorial, intermediate and apical domain. The ATP-binding site is located on the equatorial domain, which is composed of two layers of α-helices plus a number of small β-hairpins (Figure 1B) and does not contain the core layer of β-strands seen in the other folds for ATP binding [34,35].

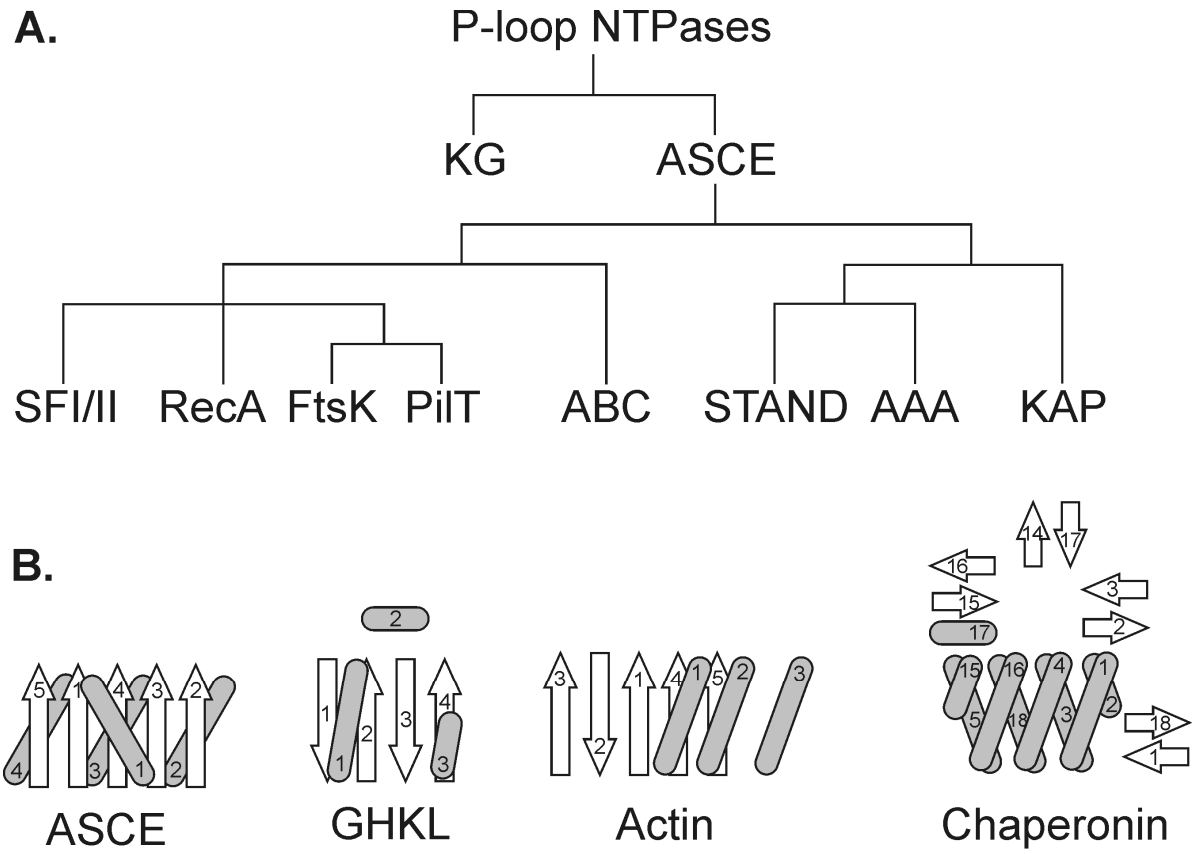


Figure 1. ATP-binding domain folds. Panel A shows the relation between the different classes of proteins within the P-loop NTPases. Panel B shows the (core) ATP-binding domains of the four different folds of ATPases [1,25,28,31,32,34]. ATP is usually bound at the interface of two domains. β-strands are depicted as white arrows and α-helices as grey cylinders, the β-strands and α-helices are numbered separately (based on the literature listed above), and numbers are not always sequential due to the fact that only the (core) nucleotide-binding domain is shown.

The class of ABC transporters

The importance of ABC transporters is not only seen in the fact that they are present in all kingdoms of life, but also by the number of diseases or phenotypes that are coupled to mutations in ABC genes i.e., diseases like cystic fibrosis, Tangier disease, Dubin

Johnson syndrome, andrenoleukodystrophy, immune deficiency and others (review [36] and the website of the human genome nomenclature committee (HGNC) website (for ABC transporters: http://nutrigene.4t.com/humanabc.htm)).

ABC transporters share a conserved overall architecture composed of a dimeric structure with two highly conserved nucleotide-binding domains (NBD) and two transmembrane domains (TMD). The two transmembrane domains form a pathway through the membrane for the translocation of the substrates. The energy needed for the translocation is provided by ATP hydrolysis by the NBDs. Export systems do not require any additional domains for their function, but most importers use an extracytoplasmic substrate-binding protein or domain (SBP for soluble periplasmic or lipid-anchored protein or SBD for domain fusions with the TMD). These SBPs/SBDs bind the substrate and deliver it to the transmembrane domain for translocation. The substrate binding domains can be expressed as a soluble protein residing in the periplasm, which is usually seen in Gram-negative bacteria. In Gram-positive bacteria, they are usually fused to the transmembrane domain or coupled to the membrane bilayer via a lipid anchor [37]. The class of energy coupling factor (ECF) transporters forms a recently discovered group of ABC-type importers (Class III importers), where the substrate-binding protein is a transmembrane domain that can couple to so-called energy-coupling modules [38]. In ABC transporters, extra domains can be present for regulation of the transporters, as seen for example in the ABC transporters CFTR [39], MalK [40] and OpuA [41,42]. The way all these domains are expressed varies widely: all domains can be expressed separately, some domains can be fused together or all domains can be fused into a single polypeptide [21].

The osmoregulated ABC transporter OpuA (*L.lactis*) has been used as a model for the class of ABC transporters throughout this thesis. The ABC transporter OpuA is involved in the uptake of the compatible solute glycine betaine to protect the bacterial cell against hyperosmotic stress. The protein contains an SBD fused to each of the TMDs, as depicted in Figure 2. The nucleotide-binding domain of this protein is fused to a C-terminal regulatory domain, the so-called cystathionine β-synthase (CBS) domains [41,42], as shown in Figure 2.

Numerous crystal structures of separate domains of ABC transporters are available, but only a few of the full ABC transporters, for instance the lipid flippase MsbA [44], the vitamin B12 transporter BtuCD [45] also in complex with its substrate binding domain BtuF [46], the maltose transporter $MalFGK_2$ [47], the multidrug exporters P-glycoprotein [48] and Sav1866 [49]. The nucleotide-binding domain is the most conserved part of ABC (transport) proteins and forms the basis for their clustering in one superfamily. A review by Schmitt and colleagues [50] in 2006 lists the nucleotide-binding domains for which crystal structures were available at that time. The structures point to a dimeric state, where the two NBDs are arranged in a head-to-tail conformation as was observed first for Rad50cd (involved in DNA repair) [51]. Several crystal structures like the *E.coli* MalK [40,52], *M.jannaschii* MJ0796 E171Q mutant [53], the *E.coli* HlyB H662A

mutant [54] and the *E.coli* BtuCD [45] underline the model for a head-to-tail structural arrangement, though other orientations have been proposed and opposed as well [55-57].

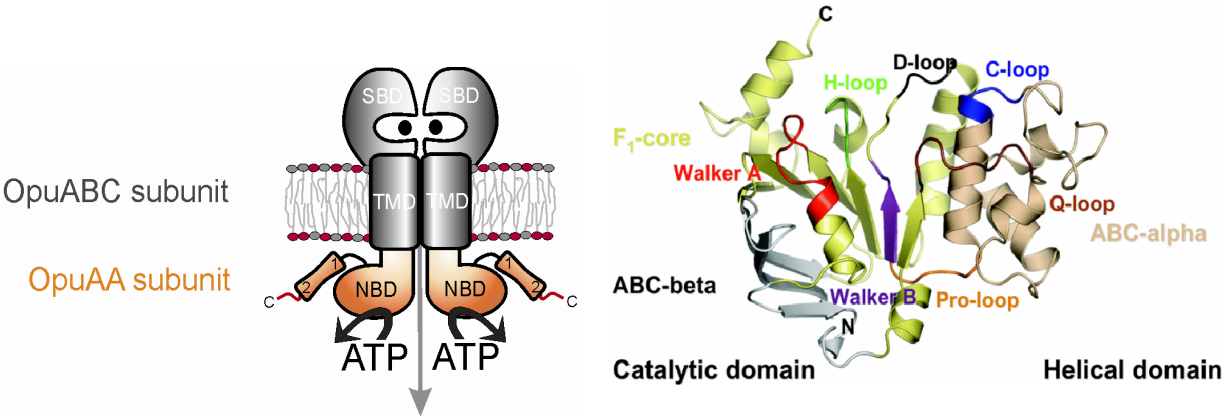


Figure 2. The left panel shows a cartoon of the ABC transporter OpuA. The substrate-binding domain (OpuAC moiety) is fused to the transmembrane domain (OpuAB moiety). The second part of the protein (OpuAA), which is expressed as a separate protein, is composed of the nucleotide-binding domain (NBD) and a regulatory domain (two so-called CBS domains, numbered 1 and 2). ABC transporters dimerize (as is shown in the picture) to form the active complex and for that reason two copies of each protein are present. The substrate glycine betaine is indicated as a black dot, bound in the SBDs, and the directionality of transport is indicated by the grey arrow. The right panel shows the structure of one ATPase domain (HlyB) with the position of the conserved motifs ([43], printed with permission from Prof. L. Schmitt).

ATP binding and hydrolysis by ABC nucleotide binding domains

The dimeric structural arrangement of the NBDs is important for the binding of the nucleotides, allowing coordination by residues from both subunits. The structure of a single nucleotide-binding domain is composed of two lobes with an overall L-shaped form. The largest domain (arm I) is called the catalytic domain and contains the majority of the conserved sides amongst which the Walker A motif (GxxGxGK[S/T], x can be any amino acid), Walker B motif (ϕϕϕϕD, ϕ can be any hydrophobic amino acid), the D-loop and the H-loop. The smaller lobe (arm II) is called the helical domain and contains the signature motif (LSGGQ). The connection between the two arms contains the conserved Q-loop and Pro-loop [50,58,59]. The different conserved motifs of the NBDs are shown in Figure 2. ATP is sandwiched between the signature motif of one NBD and the Walker A and B motifs of the second NBD. The Walker A motif interacts with the phosphates of the nucleotide and the coordination of the Mg^{2+} ion is facilitated by the β- and γ-phosphate of the nucleotide, by the side chain of the conserved serine in the Walker A motif and is further coordinated via several water molecules. One of these water molecules is bound via the Walker B motif (the aspartic acid), which also coordinates a water molecule around the γ-phosphate of ATP. The Q-loop coordinates a water molecule that interacts directly with ATP. The main interactions of the C-loop of the second NBD (*trans*-interaction) are with the γ-phosphate of ATP. The D-loop of the

second NBD interacts (*trans*-interaction) with the Walker A motif of the first NBD. In the crystal structure of HlyB, an aromatic amino acid (Y477, located on the catalytic domain) forms π-π stacking interactions with the adenine ring of ATP, which also seems to be conserved in other NBDs and might explain the preference of NBDs for adenine nucleotides. The H-loop is overlapping with residues that are essential for catalysis in other ATPases and GTPases and is proposed to function either in steps before ATP hydrolysis or to stabilize the transition state [50,54,60].

Based on structural and biochemical studies, two models for the hydrolysis cycle of the NBDs have been proposed, namely the alternating catalytic site model [61] and the processive-clamp mechanism [62]. In the alternating catalytic site model, ATP binding at the first NBD allows the ATP hydrolysis at the second site. This hydrolysis and the nucleotide release at the second site prohibits hydrolysis at the first NBD site, and only after rebinding of ATP at the second site, hydrolysis can occur again at the first nucleotide-binding site [61]. This model is based on studies using *ortho*-vanadate trapping of nucleotides in P-glycoprotein and the maltose transporter. In the maltose system, only one ADP is trapped with *ortho*-vanadate, even though ATPase activity is completely blocked under these conditions [63,64]. In the processive-clamp mechanism, two ATP molecules are bound to the subunits that form a dimeric NBD complex. After sequential hydrolysis of both ATP molecules, the dimer complex dissociates and releases ADP and inorganic phosphate [62]. This model was proposed on the basis of studies with the ABC transporter Mdl1p, where the dimeric complex was always found to contain two nucleotides (two ATPs or ATP plus ADP; two ADPs were only found after trapping with BeF_x) [62].

Outline of the thesis

The wide variety of proteins that use ATP for their function either as source of metabolic energy, building block, cofactor or signaling molecule underlines the importance of this molecule in biology. Throughout this thesis, the focus has been on one of the major classes of adenine nucleotide-binding proteins, namely the ABC transporters In order to understand more about the role of the ABC transporters in a cell, one needs information on the structure and function of the protein, and knowledge of their time- and space-resolved expression.

Chapter 2 and 3 focus on the development of a method to enrich for adenine-nucleotide binding proteins present in cell lysates. Chapter 2 describes the design and synthesis of two ATP-based probes, used as basis for the approach of activity-based protein profiling (Chapter 3). The probes were designed to fit optimally in the nucleotide-binding pocket of ABC transporters. The introduction of this chapter summarizes what is known of ATP-based nucleotide analogues. Chapter 3 presents a proof-of-principle with the probes synthesized in Chapter 2, using well-characterized ABC transport proteins.

Chapter 4 presents the characterization of the detergent-solubilized ABC transporter OpuA in respect to ligand binding (ATP and glycine betaine), which was critical for the design of the experiments presented in Chapter 3. Also, the crystal structure of the substrate-binding domain of OpuA is presented. Chapter 5 focuses on the physiological consequences of expressing ABC and other membrane proteins in *Lactococcus lactis*. This study explored another way to optimize the detection of membrane proteins in proteomics, using additional washing steps to strip the membrane from soluble and loosely-attached proteins. Chapter 5 combines the proteomics with a transcriptomics study, carried out in the Molecular Genetics group. The combined proteomics and transcriptomics study provides important information for the further engineering of *L. lactis* cells for improved membrane protein production.

References

[1] Thomsen,N.D. & Berger,J.M. (2008) Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases. *Mol Microbiol*, **69,** 1071-1090.

[2] Vale,R.D. & Milligan,R.A. (2000) The way things move: looking under the hood of molecular motor proteins. *Science*, **288,** 88-95.

[3] Kardon,J.R. & Vale,R.D. (2009) Regulators of the cytoplasmic dynein motor. *Nat. Rev Mol Cell Biol*, **10,** 854-865.

[4] Neuwald,A.F., Aravind,L., Spouge,J.L., & Koonin,E.V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res*, **9,** 27-43.

[5] Kolberg,M., Strand,K.R., Graff,P., & Andersson,K.K. (2004) Structure, function, and mechanism of ribonucleotide reductases. *Biochim. Biophys Acta*, **1699,** 1-34.

[6] Dvir,A. (2002) Promoter escape by RNA polymerase II. *Biochim. Biophys Acta*, **1577,** 208-223.

[7] Richardson,J.P. (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim. Biophys Acta*, **1577,** 251-260.

[8] Davey,M.J., Jeruzalmi,D., Kuriyan,J., & O'Donnell,M. (2002) Motors and switches: AAA+ machines within the replisome. *Nat. Rev Mol Cell Biol*, **3,** 826-835.

[9] Bloom,L.B. (2006) Dynamics of loading the Escherichia coli DNA polymerase processivity clamp. *Crit Rev Biochem Mol Biol*, **41,** 179-208.

[10] Hall,M.C. & Matson,S.W. (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol*, **34,** 867-877.

[11] Bates,A.D. & Maxwell,A. (2007) Energy coupling in type II topoisomerases: why do they hydrolyze ATP? *Biochemistry*, **46,** 7929-7941.

[12] Cox,M.M. (2003) The bacterial RecA protein as a motor protein. *Annu. Rev Microbiol*, **57,** 551-577.

[13] Moras,D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem Sci*, **17,** 159-164.

[14] Woriax,V.L., Spremulli,G.H., & Spremulli,L.L. (1996) Nucleotide and aminoacyl-tRNA specificity of the mammalian mitochondrial elongation factor EF-Tu.Ts complex. *Biochim. Biophys Acta*, **1307,** 66-72.

[15] Chakraburtty,K. (1999) Functional interaction of yeast elongation factor 3 with yeast ribosomes. *Int J Biochem Cell Biol*, **31,** 163-173.
[16] Harden,T.K., Boyer,J.L., & Nicholas,R.A. (1995) P2-purinergic receptors: subtype-associated signaling responses and structure. *Annu. Rev Pharmacol Toxicol.*, **35,** 541-579.
[17] Pedersen,P.L. (2007) Transport ATPases into the year 2008: a brief overview related to types, structures, functions and roles in health and disease. *J Bioenerg. Biomembr.*, **39,** 349-355.
[18] Pedersen,P.L. & Carafoli,E. (1987) Ion Motive Atpases .2. Energy Coupling and Work Output. *Trends in Biochemical Sciences*, **12,** 186-189.
[19] Nelson,N. (1992) The vacuolar H(+)-ATPase--one of the most fundamental ion pumps in nature. *J Exp. Biol*, **172,** 19-27.
[20] Davidson,A.L., Dassa,E., Orelle,C., & Chen,J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev*, **72,** 317-64, table.
[21] Biemans-Oldehinkel,E., Doeven,M.K., & Poolman,B. (2006) ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett*, **580,** 1023-1035.
[22] Saraste,M., Sibbald,P.R., & Wittinghofer,A. (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci*, **15,** 430-434.
[23] Walker,J.E., Saraste,M., Runswick,M.J., & Gay,N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J*, **1,** 945-951.
[24] Aravind,L., Iyer,L.M., Leipe,D.D., & Koonin,E.V. (2004) A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. *Genome Biol*, **5,** R30.
[25] Iyer,L.M., Makarova,K.S., Koonin,E.V., & Aravind,L. (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res*, **32,** 5260-5279.
[26] Herdendorf,T.J., McCaslin,D.R., & Forest,K.T. (2002) Aquifex aeolicus PilT, homologue of a surface motility protein, is a thermostable oligomeric NTPase. *J Bacteriol*, **184,** 6465-6471.
[27] Iyer,L.M., Leipe,D.D., Koonin,E.V., & Aravind,L. (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol*, **146,** 11-31.
[28] Erzberger,J.P. & Berger,J.M. (2006) Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev Biophys Biomol. Struct*, **35,** 93-114.
[29] Danot,O., Marquenet,E., Vidal-Ingigliardi,D., & Richet,E. (2009) Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. *Structure*, **17,** 172-182.
[30] Leipe,D.D., Koonin,E.V., & Aravind,L. (2004) STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol*, **343,** 1-28.
[31] Dutta,R. & Inouye,M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*, **25,** 24-28.

[32] Bork,P., Sander,C., & Valencia,A. (1992) An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl. Acad. Sci U. S. A*, **89,** 7290-7294.
[33] Kabsch,W. & Holmes,K.C. (1995) The actin fold. *FASEB J*, **9,** 167-174.
[34] Boisvert,D.C., Wang,J., Otwinowski,Z., Horwich,A.L., & Sigler,P.B. (1996) The 2.4 A crystal structure of the bacterial chaperonin GroEL complexed with ATP gamma S. *Nat. Struct Biol*, **3,** 170-177.
[35] Braig,K., Otwinowski,Z., Hegde,R., Boisvert,D.C., Joachimiak,A., Horwich,A.L., & Sigler,P.B. (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature*, **371,** 578-586.
[36] Dean,M. (2005) The genetics of ATP-binding cassette transporters. *Methods Enzymol*, **400,** 409-429.
[37] Biemans-Oldehinkel,E., Doeven,M.K., & Poolman,B. (2006) ABC transporter architecture and regulatory roles of accessory domains. *Febs Letters*, **580,** 1023-1035.
[38] Rodionov,D.A., Hebbeln,P., Eudes,A., ter Beek,J., Rodionova,I.A., Erkens,G.B., Slotboom,D.J., Gelfand,M.S., Osterman,A.L., Hanson,A.D., & Eitinger,T. (2009) A novel class of modular transporters for vitamins in prokaryotes. *J Bacteriol*, **191,** 42-51.
[39] Cheung,J.C., Kim,C.P., Pasyk,S., & Bear,C.E. (2008) Molecular basis for the ATPase activity of CFTR. *Arch Biochem Biophys*, **476,** 95-100.
[40] Chen,J., Lu,G., Lin,J., Davidson,A.L., & Quiocho,F.A. (2003) A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol Cell*, **12,** 651-661.
[41] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. USA*, **103,** 10624-10629.
[42] Mahmood,N.A., Biemans-Oldehinkel,E., & Poolman,B. (2009) Engineering of ion sensing by the cystathionine beta-synthase module of the ABC transporter OpuA. *J Biol Chem*, **284,** 14368-14376.
[43] Zaitseva,J., Jenewein,S., Oswald,C., Jumpertz,T., Holland,I.B., & Schmitt,L. (2005) A molecular understanding of the catalytic cycle of the nucleotide-binding domain of the ABC transporter HlyB. *Biochem Soc Trans*, **33,** 990-995.
[44] Ward,A., Reyes,C.L., Yu,J., Roth,C.B., & Chang,G. (2007) Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl. Acad. Sci U. S. A*, **104,** 19005-19010.
[45] Locher,K.P., Lee,A.T., & Rees,D.C. (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, **296,** 1091-1098.
[46] Hvorup,R.N., Goetz,B.A., Niederer,M., Hollenstein,K., Perozo,E., & Locher,K.P. (2007) Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. *Science*, **317,** 1387-1390.
[47] Oldham,M.L., Khare,D., Quiocho,F.A., Davidson,A.L., & Chen,J. (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*, **450,** 515-521.
[48] Aller,S.G., Yu,J., Ward,A., Weng,Y., Chittaboina,S., Zhuo,R., Harrell,P.M., Trinh,Y.T., Zhang,Q., Urbatsch,I.L., & Chang,G. (2009) Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, **323,** 1718-1722.

[49] Dawson,R.J. & Locher,K.P. (2007) Structure of the multidrug ABC transporter Sav1866 from Staphylococcus aureus in complex with AMP-PNP. *FEBS Lett*, **581,** 935-938.

[50] Oswald,C., Holland,I.B., & Schmitt,L. (2006) The motor domains of ABC-transporters. What can structures tell us? *Naunyn Schmiedebergs Arch Pharmacol*, **372,** 385-399.

[51] Hopfner,K.P., Karcher,A., Shin,D.S., Craig,L., Arthur,L.M., Carney,J.P., & Tainer,J.A. (2000) Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell*, **101,** 789-800.

[52] Lu,G., Westbrooks,J.M., Davidson,A.L., & Chen,J. (2005) ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation. *Proc. Natl. Acad. Sci U. S. A*, **102,** 17969-17974.

[53] Smith,P.C., Karpowich,N., Millen,L., Moody,J.E., Rosen,J., Thomas,P.J., & Hunt,J.F. (2002) ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell*, **10,** 139-149.

[54] Zaitseva,J., Jenewein,S., Jumpertz,T., Holland,I.B., & Schmitt,L. (2005) H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J*, **24,** 1901-1910.

[55] Diederichs,K., Diez,J., Greller,G., Muller,C., Breed,J., Schnell,C., Vonrhein,C., Boos,W., & Welte,W. (2000) Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon Thermococcus litoralis. *EMBO J*, **19,** 5951-5961.

[56] Hung,L.W., Wang,I.X., Nikaido,K., Liu,P.Q., Ames,G.F., & Kim,S.H. (1998) Crystal structure of the ATP-binding subunit of an ABC transporter. *Nature*, **396,** 703-707.

[57] Yuan,Y.R., Blecker,S., Martsinkevich,O., Millen,L., Thomas,P.J., & Hunt,J.F. (2001) The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *J Biol Chem*, **276,** 32313-32321.

[58] Linton,K.J. (2007) Structure and function of ABC transporters. *Physiology (Bethesda.)*, **22,** 122-130.

[59] Schmitt,L., Benabdelhak,H., Blight,M.A., Holland,I.B., & Stubbs,M.T. (2003) Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. *J Mol Biol*, **330,** 333-342.

[60] Zaitseva,J., Oswald,C., Jumpertz,T., Jenewein,S., Wiedenmann,A., Holland,I.B., & Schmitt,L. (2006) A structural analysis of asymmetry required for catalytic activity of an ABC-ATPase domain dimer. *EMBO J*, **25,** 3432-3443.

[61] Senior,A.E., al Shawi,M.K., & Urbatsch,I.L. (1995) The catalytic cycle of P-glycoprotein. *FEBS Lett*, **377,** 285-289.

[62] Janas,E., Hofacker,M., Chen,M., Gompf,S., van der,D.C., & Tampe,R. (2003) The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p. *J Biol Chem*, **278,** 26862-26869.

[63] Sharma,S. & Davidson,A.L. (2000) Vanadate-induced trapping of nucleotides by purified maltose transport complex requires ATP hydrolysis. *J Bacteriol*, **182,** 6570-6576.

[64] Urbatsch,I.L., Sankaran,B., Weber,J., & Senior,A.E. (1995) P-glycoprotein is stably inhibited by vanadate-induced trapping of nucleotide at a single catalytic site. *J Biol Chem*, **270,** 19383-19390.

Chapter 2 Design and synthesis of ATP-based nucleotide analogues for the profiling of adenine-nucleotide binding proteins

Justina. C. Wolters, Gerard Roelfes and Bert Poolman

Abstract

Two nucleotide-based probes were designed and synthesized in order to enrich samples for specific classes of proteins by affinity-based protein profiling. The idea is to use the probes (the baits) to enrich specific classes of proteins from complex mixtures. We focused on the profiling of adenine nucleotide-binding proteins. Two properties were considered in the development of the probes: the bait needs to bind adenine nucleotide-binding proteins with high affinity and carry a second functional group suitable *and well accessible* for coupling to a chromatography resin. For this purpose, we synthesized *p*-biotinyl amidobenzoic acid-ATP (*p*-BABA-ATP) and *p*-biotinyl aminomethylbenzoic acid-ATP (*p*-BAMBA-ATP). The structure of the products was established by nuclear magnetic resonance and mass spectrometry based methods. *p*-BABA-ATP and *p*-BAMBA-ATP both have a binding affinity at least 10-fold higher compared to unmodified ATP.

Introduction

For the optimal design of the nucleotide analogues, extensive screening was done to gain some understanding of the adenine nucleotide-binding pocket and its bound nucleotides. In the first approach, the nucleotide-binding pocket itself was analyzed in crystal structures, present in the protein databank (PDB). These different crystal structures yielded information about the accessibility of the nucleotide and the amount of space in the binding pocket to allow for modifications on the nucleotide. In the second approach, an overview was made of previously synthesized ATP analogues. The combination of both searches yielded a strategy for the design and synthesis of novel probes to profile (adenine) nucleotide-binding proteins, with a focus on ABC transporters.

Adenine nucleotide-binding proteins

The analysis of crystal structures was done to gain insight into the space in and around the nucleotide-binding pocket to fit in nucleotide analogues. Binding of the analogues can be blocked by steric hindrance and disturbed interactions between the nucleotide and the protein. Although it is difficult to make predictions about the interactions on the basis of crystal structures, the analysis of the binding pocket can yield valuable information about the chance that steric hindrance will occur upon addition of bulky groups. The X-ray crystallography structures of various proteins with nucleotides bound

were examined and their binding pockets estimated with the program Q-SiteFinder [1], as shown in Supplement I. The crystal structures were grouped according to the four main folds (introduced in Chapter 1) and crystal structures that were not explicitly grouped in one of these four main classes are listed separately.

From the nucleotide-binding proteins with the ASCE fold, the class of ABC transporters has the tightest binding pocket. The space around the phosphate and nucleobase of the ATP limits modifications on these positions, but the ribose moiety is close to an open cavity (seen clearly in the space-filling models of the crystal structure, Supplement I). In the FtsK, pilT, AAA- and F-ATPases, the ribose moiety is not even modelled in the binding pocket itself, but is already in this cavity allowing even better access for modifications. In addition to the accessibility of the ribose moiety, the nucleobase in FtsK, PilT and RecA is also part of a larger pocket or in an open cavity, potentially allowing modifications of the nucleobase as well.

Proteins with the GHKL and actin fold do not show clearly overlapping patterns in regard to space in the nucleotide-binding pocket. For instance, in MutL the adenine ring is in an open cavity, in the heat shock protein Hsp90 the ribose is in an open cavity with the phosphate group close by, and in the gyrase GyrB the whole nucleotide is enclosed without a clear access point to a cavity. In the actin fold proteins a pattern might have evolved for accessibility of the phosphate groups (open cavity for Hsc70 and space in the binding pocket for actin and the hexokinase), whereas the nucleobase is nearby the surface at two of the proteins.

Based on the crystal structure of the chaperonin GroEL, modifications at any of the ATP moieties will be allowed. This spacing around the nucleotides is also observed in the P-loop NTPase kinases, RNA polymerase II, the ribonucleotide reductase and glycyl tRNA synthetase, while the adenylyl cyclase nucleotide is only restricted at the nucleobase moiety. Even though there seems to be no spatial restriction upon modification of these nucleotides, one should keep in mind that specific interactions between the analogue and the binding pocket may prevent the nucleotide to bind efficiently.

The binding pocket of the elongation factor Tu shows a tight pocket, with accessibility of the ribose moiety of ATP similar to that in ABC transporters. The motor ATPases myosin and kinesin also have the ribose moiety in an open cavity. The P-type ATPases show the opposite behaviour compared to the ABC transporters. In the crystal structures, the ribose moiety seems least accessible and the phosphate groups most accessible.

ATP analogues

Screening of literature shows a large number of ATP analogues, which can be subdivided in three classes based on modifications in one of the three functional groups: (i) the nucleobase (adenine ring); (ii) the sugar group (ribose ring); and (iii) the triphosphate moiety. Within each of these classes a distinction can be made in terms of added

functionality: analogues that act as reporters or analogues that lock a protein in a specific state either via non-covalent binding (like non-hydrolyzable analogues) or covalent cross-linking. The group of reporter analogues can be divided in molecules with fluorescent (or other NMR or EPR-specific functionalities) or radio-active labeled moieties and probes that can bind to antibodies or column based materials (like a biotin-moiety). Excellent reviews can be found in [2-5]; for an overview see Supplement II, Table S1 (the analogues described in the supplement table are listed in Table 1). The more recent reviews by Jameson [4] and Cremo [3] discuss the characteristics and general methods to synthesize ATP analogues, with a focus on fluorescent derivatives. The review by Yount [5] is the most comprehensive in terms of classes of ATP analogues; the paper by Bagshaw [2] gives a nice visual overview of the modifications at the different reactive moieties of ATP. Supplement II, Table S1, presents the ATP analogues and the protein classes targeted by these analogues; Table S2 presents the (commercially) available resin-bound nucleotides and other types of molecules to enrich for adenine nucleotide-binding proteins. Below, we summarize the most salient features of the different classes of analogues.

Table 1. Overview of the different ATP analogues that are listed in Supplement II, Table S1, sorted according their function and modification site (abbreviations are explained in the supplementary table).

Functionality	**Nucleobase-modified**	**Phosphate-modified**	**Ribose-modified**
Fluorophore	εATP *lin*-Benzo-ATP MABA-ATP	AmNS-ATP EDANS-ATP Pγ-MABA-ATP	DANS-ATP DAN-ATP (M)ANT-ATP TNP-ATP All EDA-ATP analogues FDNP-ATP
Non-hydrolyzable	-	ATPαS, ATPβS, ATPγS AMPCP-P, AMP-PCP AMP-PNP	-
Crosslinking	(2- and 8-) Azido-ATP	ATPγBP Azidoanilide-ATP ClRATP	oATP BzATP Arylazido-ATP FDNP-ATP
Biotinylated	Bio-17-ATP	Azido-ATP(γ)-biotin ATP(γ)-biotin-pentylamine ATP(γ)-biotin-LC-PEO-amine	Azido-biotinyl-ATP
Spectroscopic	NSL, (am)TEMPO-ATP		-
Other	2-amino-ATP 8-Br-ATP 6/8-alkylated-ATP	ATP-α-B ATPγF	2’ or 3’-methyl-ATP $_N$ATP Ara-ATP, xylo-ATP 2-Cl-ATP

Fluorescent ATP analogues

Coupling of different fluorescent moieties was achieved on all three functional groups of ATP, as listed in Table 1 (and Supplement II, Table S1). The fluorescent (methyl)anthroyl-group was coupled to all three sides of ATP, namely to the nucleobase (MABA-ATP), the phosphates (Pγ-MABA-ATP) and the ribose moiety (MANT-ATP). A clear difference can

be observed in their application: the phosphate and nucleobase analogues have only been used for heat shock proteins (chaperones), while the ribose-modified nucleotides have been used (amongst others) to investigate kinases, adenylyl cyclase, a P-type ATPase, and the molecular motor proteins myosin, dynein and kinesin. Modifications with sulfonate-naphthalene-groups have been coupled to the phosphates (AmNS-ATP and EDANS-ATP) and ribose (DANS-ATP) moieties of ATP. The phosphate-coupled derivatives have been used as substrates for RNA polymerase. The ribose-modified derivatives have been used to study F-ATPase (DNS-ATP) and Na^+/K^+-(P-type) ATPase (DANS-ATP). The ethyldiamine (EDA)-linker, used in EDANS-ATP to couple the sulfonate-naphtalene-group to the phosphate moiety of ATP, has been used for the coupling of a large number of other fluorophores (dansyl, rhodamine, fluorescein, cynanine, coumarin, BODIPY, alexa) to the ribose moiety of ATP. Two other analogues are based on nitrophenyl-moieties (TNP-ATP and FDNP-ATP). TNP-ATP has been used to probe nucleotide binding in all different types of adenine nucleotide-binding proteins. The FDNP-ATP analog has not been used extensively despite its dual functionality as fluorophore and crosslinker.

Two modifications specific for the nucleobase comprise an additional ring structure, yielding εATP and lin-Benzo-ATP. These modifications yield fluorophores with relatively small changes of the overall structure of the nucleotide. In *lin*-Benzo-ATP, an extra ring is present in between the two-ring structure of the adenine moiety, which yields a stretched-out molecule. The analog can be used to screen the space in the binding pocket of adenine nucleotide-binding proteins, like it was done for a number of kinases and transport ATPases. However, the modification was not tolerated by proteins like the RNA polymerase or the H^+/K^+- (P-type) ATPase, even though the Mg^{2+}- (P-type) ATPase tolerated this probe [6,7]. Overall, the fluorescent ribose-modified ATP analogues have seen many applications with TNP-ATP and MANT-ATP being the most popular. The predominant choice of ribose-modified nucleotide analogues matches well with the spacing observed in the crystal structures of many adenine nucleotide-binding proteins.

Non-hydrolyzable ATP analogues

The non-hydrolyzable ATP analogues may lock nucleotide-binding proteins in a specific state. The modifications are in the phosphate moiety to prevent hydrolysis to occur, which can be done by replacing bridging or non-bridging oxygens by carbons, sulfur or nitrogen (for a review see Young 1975 [5]). The annotation of these analogues as 'non-hydrolyzable' should be taken with caution because, depending on the enzyme, they may be slowly hydrolyzed. The use of these ATP analogues as (competitive) inhibitors and probes to lock the protein in a specific sub-state has been extensively exploited. When the normal P-O-P bond is replaced by P-N-P (AMP-PNP) or P-C-P (AMP-PCP and AMPCP-P), the bond length and the bond angles change [5]. Of the analogues AMP-PNP is most similar to ATP and used more often than AMP-PCP or AMPCP-P analogues. AMP-

PCP has been used to study purinergic receptors and P-type ATPases, while AMP-PNP has been used in studies of kinases, myosin, kinesin, F-ATPase, adenylyl cyclase, topoisomerase, and ABC transporters like CFTR. The modifications of non-bridging oxygens in the thiophosphate analogues (ATPαS, ATPβS and ATPγS) do not alter the bond length or bond angle relative to the triphosphates. Of these three thiophosphate analogues, ATPγS has been used most extensively. The ATPαS and ATPβS have been used less frequently and are more sensitive to hydrolysis because of the unmodified γ-phosphate. The presence of two diastereomers [8] in these analogues has been used to characterize the specificity of proteins.

Crosslinking

Covalent crosslinking of ATP analogues to proteins can be done either via chemical affinity labeling or photoaffinity labeling. A previous review by Colman in 1983 [9] describes the use of several affinity labels in purine nucleotide sites, but also elaborates widely on the use of several adenosine analogues, which are omitted from our Supplement II, Table S1. Crosslinking of the ATP analogues to the binding pocket can be used to study the binding of nucleotides in detail, their (apparent) affinity as well as binding stoichiometries, but characterization of the amino acids that are crosslinked to the nucleotide analog also enables analysis of the nucleotide-binding pocket itself. The azido (N_3) group crosslinks non-specifically upon photo-activation via formation of a nitrene. Coupling of the N_3 group to ATP will direct labeling to the nucleotide-binding pocket of proteins, while unspecific binding is lowered by fast reaction of the azido-group with the solvent. ATP analogues have been prepared with the azido-group on all three groups: the nucleobase (2-azido- and 8-azido-ATP), the phosphates (azidoanilide-ATP) and the ribose (arylazido-ATP). A clear difference can be observed in the use of the different analogues. The ribose-modified nucleotide has been used to label the binding-pocket or study the binding affinity/stoichiometry of P- and F-type ATPases and myosin, while the phosphate modified analogues have been used for kinases and synthetases. The nucleobase-modified analogues have been used for practically all adenine nucleotide-binding proteins. The 2-azido-ATP analog has been used most for F-type ATPases, while 8-azido-ATP has been used more frequently for ABC (multidrug) transporters. The difference between 2-azido-ATP and 8-azido-ATP is based on the different conformation they adopt: substitution of the 8-position induces a *syn*-conformation of the ATP analog, while substitution of the 2-position retains the ATP analog in an *anti*-conformation, most likely due to steric hindrance in the *syn*-form. The analysis and interpretation of experiments performed with the 2-azido-ATP though is complicated by the formation of non-photoreactive tautomers (tetrazolo isomers) [10]. The percentage of these tautomers in the reaction is unpredictable, because formation is highly dependent on the experimental conditions. These side effects might explain the preference for 8-azido-ATP over 2-azido-ATP as observed in the literature, even though

2-azido-ATP seems more effective for proteins like the F-type ATPase, dynein and myosin.

The second type of photo-crosslinkers, the benzophenone (benzoyl)-derivatives, belong to the class of α,β-unsaturated carbonyl compounds that crosslink via photochemically generated free radicals (or triplet states) [11]. The reactive group was attached either to the phosphate (ATPγBP) or the ribose moiety (BzATP). The phosphate-modified analog has been used for the characterization of the binding pocket of two kinases and a carboxylase, while the ribose-modified analog has been used more extensively, not only for kinases but also for molecular motors like myosin and kinesin, transport ATPases like F- and V-ATPases and even for purinergic receptors. The same benzoyl-group has been deployed for ATP mimics like p-fluorosulfonylbenzoyl adenosine (3'-FSBA/5'-FSBA) [12,13] and 4-benzoyl(benzoyl)-1-amidofluorescein (BzAF) [14] as is discussed briefly under 'Other ATP analogues'.

The non-photoactivatable crosslinkers like the ribose-modified periodate-ATP (oATP) and FDNP-ATP, and the phosphate-modified ClRATP are used less often. In the ribose-modified (oATP), the oxidized 2',3'-aldehyde can covalently link with the ε-amino side chain of lysine in the presence of sodium borohydride [15]. Periodate-oxidized nucleotides have been used to study purinergic receptors and P-type ATPases. Similar to the nucleobase modified azido-ATP analogue, oATP has the advantage that the modification does not involve bulky groups for the crosslinking. The ribose modified FDNP-ATP contains the weakly reactive 3'-O-5-fluoro-2,4-dinitrophenyl (FDNP)-moiety which can react *in situ* with lysines, tyrosines or cysteines (forming protein-DNP-ATP complexes). This analog has been used to study F- and P-type ATPases, an uncoupling protein and a caspase (cysteine protease). The phosphate-modified ClRATP forms an ester bond with the protein upon crosslinking and has been used to label amino acids in the active site of a Na^+/K^+- (P-type) ATPase and creatine kinase.

Overall, the azido-group has been most widely as the modification is relatively small in these ATP analogues. The additional advantage is the ease to combine this modification with other types of modifications (see Supplement II, Table S1). The benzoyl-moiety is much larger than the azido-group, however, this group shows a preference for crosslinking with carbon-hydrogen bonds over oxygen-hydrogen bonds (from the solvent), yielding higher crosslinking efficiencies, compared to azido-groups. The reduced labelling efficiency of azide derivatives upon crosslinking with solvent is counterbalanced by the advantage that the reaction with the solvent reduces the unspecific labelling of protein samples [11].

Biotinylated ATP analogues

The biotin-group can be used to bind to streptavidin-based column materials as well as applications involving detection via antibodies. Modification with biotin adds only a relative small functional group to the nucleotide, while the affinity of the biotin to streptavidin is one of the highest affinities reported in nature (K_D of 10^{-15} M for biotin

avidin binding [16]). The coupling of the biotin has been achieved at the nucleobase (Bio-17-ATP), the ribose (azido-biotinyl-ATP) and the phosphate (ATPγ-biotin, with various linkers), but the analogues have not been widely used for the characterization of adenine nucleotide-binding proteins. The nucleobase-modified analog has been used in *in vitro* transcription reactions. One of the phosphate-modified variants has been used in the analysis of nucleotide-binding to immunoglobulins and a synaptic vesicle protein. The ribose-modified analog has been used for the detection of V- and F-ATPases and a guanylyl cyclase receptor.

Spin-labeled ATP analogues

Spin-labeled (ATP) analogues are used to study the mobility and conformations of adenine nucleotide-binding sites via the stable free radical in the spin label, using the electron paramagnetic resonance (EPR) technique. The location of the spin label is important for the sensitivity of the analog and up to date the spin label has been only introduced in the nucleobase moiety of ATP [NSL and (am)TEMPO-ATP]. The spin labels have been used to probe actomyosin, creatine kinase, glutamine synthetase and RNA polymerase.

Other analogues

Various other types of nucleotide analogues have been synthesized, like alkylated and halogenated derivatives and nucleotide-containing sugar-moieties other than ribose. Nucleobase-alkylated ATP analogues, in which the conformation (syn/anti) of 8-substituted ATP was varied, have been prepared by Takenaka [17] and were used to characterize nucleotide binding to (acto)myosin. The ribose-modified alkylated-ATP analog (methyl-ATP) has been used to screen the importance of hydroxyl-groups on the ribose ring for binding to (acto)myosin. Halogenated ATP analogues were prepared with the functional group on all three moieties of ATP: the nucleobase (8-Br-ATP), the ribose (2-Cl-ATP) and the phosphate moiety (ATPγF). These molecules have been used to characterize adenine nucleotide-binding proteins and served as basis for further modifications. The fluoro-ATP can be also used in nuclear magnetic resonance (NMR) studies.

Chromatography resin-bound ATP (analogues) are listed in Supplement II, Table S2. Most of them are based on the ATP analogues described heretofore; see also Supplement II, Table S1. The linkers provide spacing between the nucleotide-binding proteins and the resin materials; the choice of the type of column (which ATP analog is used and how it is linked to the resin) is dependent on the side where the specific nucleotide-binding proteins tolerate the modifications best.

ATP mimics like 2-(dansylamino)ethyl nucleotides, *p*-fluorosulfonylbenzoyl adenosine and 4-benzoyl(benzoyl)-1-amidofluorescein are not listed in Supplement II, Table SI. In the 2-(dansylamino)ethyl triphosphate (DTP), the ethanolamine-group mimics the ribose ring in ATP and the adenine ring is mimicked by the dansyl-group [18]. The *p*-fluorosulfonylbenzoyl adenosine (5'-FSBA) is a more general nucleotide derivative

that can substitute both ATP and NADH [12,19], whereas 4-benzoyl(benzoyl)-1-amidofluorescein (BzAF) is a bifunctional reporter with a fluorescent (fluorescein) and a crosslinking moiety (benzoyl-group) [14]. Adenine or adenosine derivatives (often used for drug related purposes) are excluded from Supplement II, Table S1 even though these derivates could be used as a building block for the synthesis of ATP.

General patterns that can be seen from these analyses is that small modifications at the nucleobase (like azido-ATP, εATP) and phosphate (AMP-PNP, AMP-PCP and ATPγS) of ATP are tolerated by most adenine nucleotide-binding proteins, but the use of analogues containing bulkier groups at these positions are restricted more to the classes of kinases and DNA/RNA-modifying proteins. The crystal structures of the nucleotide-binding pockets of kinases and DNA/RNA related proteins like the RNA polymerase already indicated that there might be space around the phosphate and/or nucleobase moiety of ATP to harbor additional functionalities.

The classes of ASCE fold proteins containing the V-, F- and ABC- transport ATPases have been investigated most extensively with ribose-modified nucleotides, consistent with our predictions on basis of the crystal structures. A limitation of the analysis of crystal structures is seen in the other types of transport ATPases, the P-type ATPases. This class of transporters was studied mostly with ribose-modified nucleotides, while the space around the ribose moiety in the nucleotide binding domain crystal structure seems restricted. The majority of the overview presented, nevertheless, suggests that both size and position of the modifications to the nucleotides can be optimized for specific classes of proteins and this knowledge was used to design two nucleotide analogues for the profiling of adenine nucleotide-binding proteins of the ABC superfamily. Here, we describe the synthesis and characterization of ATP analogues with three functionalities: the nucleotide, a linker moiety and biotin, with the biotinylated linker attached to the ribose moiety of the nucleotide.

Results and Discussion

Design of the ATP analogues

The ATP analogues were designed with three functionalities: the nucleotide (ATP), a linker moiety and biotin. The analogues were designed for the profiling of ABC transporters and the modifications were made to the ribose ring. The linker should keep the biotin group accessible for binding to a chromatography resin. A second purpose of the linker is to increase the affinity of proteins for the ATP analog compared to normal ATP. Previous results obtained with 8-N_3-3'-biotinyl-ATP [20] and TNP-ATP [21] served as basis for the design of the analogues. 8-N_3-3'-biotinyl-ATP [20] has been used to label F-type ATPases via covalent crosslinking, and the biotin group has been used for immunodetection. Inspection of the crystal structure of the nucleotide-binding domains of the ABC transporter MJ0796 (1L2T, Figure 1A) [22] suggested that a linker would be needed to keep the biotin accessible for binding to streptavidin (see also Figure 1B). The purpose of the linker region to increase the affinity of the nucleotide analog compared

to ATP is important, because ABC transporters have been reported to bind ATP only in submillimolar to low millimolar range, which is too low for efficient profiling of the proteins. For the fluorescent ATP analog TNP-ATP, it has been observed that the affinity of binding to nucleotide-binding proteins of ABC transporters can be 2- to 3-orders of magnitude higher than for ATP [23]. The high affinity might come from the aromatic moiety and/or from the nitro-groups. Guided by these observations, the aromatic ring-structure was used as basis for the design of the linkers to couple the biotin moiety to ATP. Here, we describe the synthesis of two adenine-nucleotide analogues, biotinylamidobenzoic acid-ATP (*p*-BABA-ATP) and biotinylamidomethyl benzoic acid-ATP (*p*-BAMBA-ATP), their structural characterization and their binding to the ABC transporter OpuA.

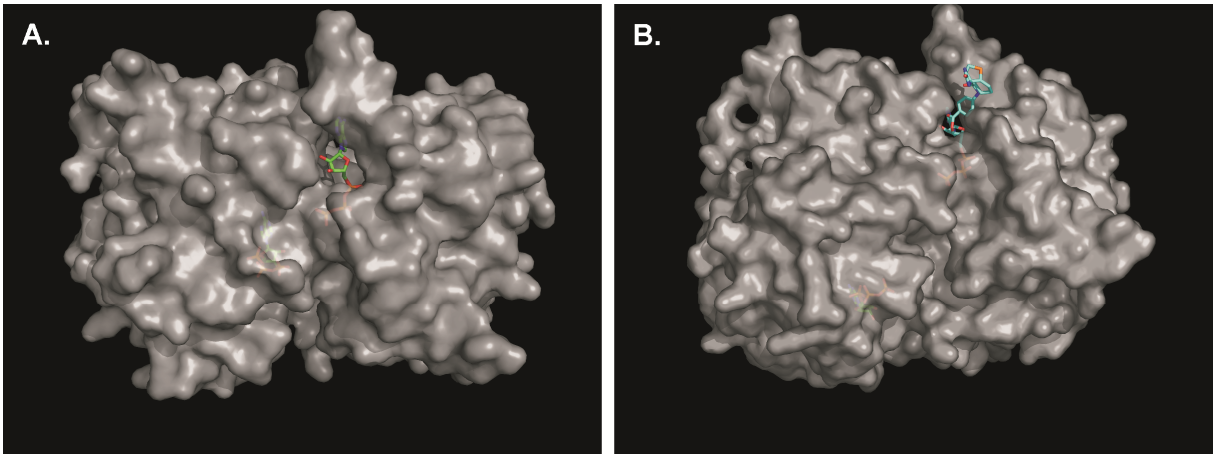


Figure 1. X-ray crystallography structure of the ABC transporter MJ0796 (1L2T) [22]. The structure was solved in the presence of ATP (panel A, ATP green), showing the space around the ribose moiety. The nucleotide was overlayed with the ATP analog *p*-BABA-ATP (panel B, ATP analog in blue) to show the accessibility of the biotin group.

Synthesis of the ATP analogues

The synthesis workflow was tested and optimized for biotinyl-amidobenzoic acid-ATP (*p*-BABA-ATP). The synthesis of biotinyl-amidobenzoic acid (Figure 2A) was done as described previously for biotinylated amino acid precursors, using isobutyl chloroformate to activate the biotin prior to coupling to the amine of the linker moiety [24]. Different water/ethanol ratios were tested to optimize the recrystallisation conditions of the synthesis. Higher ethanol concentrations facilitated the crystallization of the biotinylated linker although a small amount of water helped to remove impurities as seen in the low ppm-region of the ^{1}H-NMR-spectrum. The ^{1}H and ^{13}C-NMR-spectra (Supplement III) of the synthesized compound were comparable to the spectra published by Skander et al. 2004 [24]. A second linker (4-aminomethylbenzoic acid) was used to synthesize 4-biotinylamidomethylbenzoic acid (*p*-BAMBA) (Figure 2A), which was less soluble in DMF compared to biotinyl-amidobenzoic acid. The ^{1}H-NMR spectrum of the recrystallized *p*-BAMBA (Supplement III) was assigned on the basis of the spectra of the starting-materials and the expected shifts based on the coupling of biotin to the linker. The coupling of the biotinylated linker to adenosine triphosphate (ATP) was based

on a method described for 3'-O-(4-benzoyl)benzoyl-ATP [25]. This method involves the activation of the acid-group from biotin with carbonyldiimidazole [26], which subsequently can react with a hydroxyl group to form an ester on the ribose ring of ATP (Figure 2B).

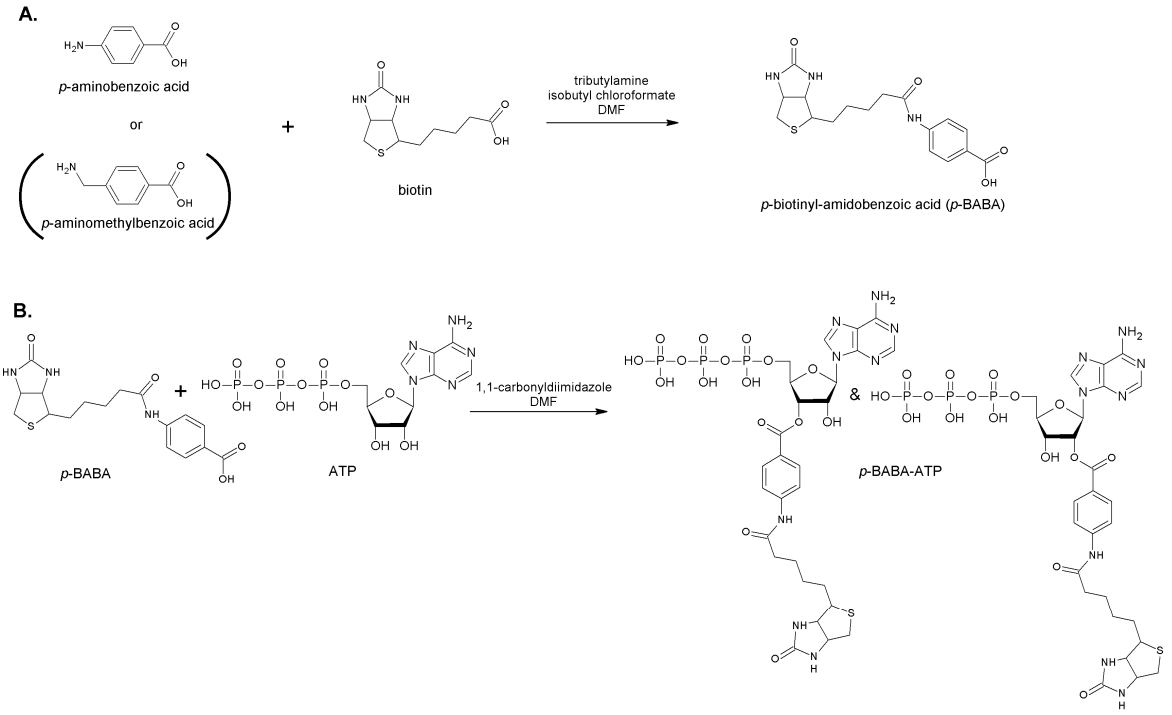


Figure 2. Chemical structures and reaction schemes for the synthesis of the two ATP analogues. Panel A: Synthesis of the biotinylated linkers *p*-biotinyl-amidobenzoic acid (*p*-BABA) and p-biotinyl-amidomethylbenzoic acid (*p*-BAMBA). Panel B: Synthesis of *p*-biotinyl-amidobenzoic acid-ATP, shown in two isoforms, from the biotinylated linker and ATP.

Optimization of the synthesis

The synthesis was optimized in terms of the ratios of the two starting materials and the reaction time. The ratio between ATP and *p*-BABA was varied from 1:8 to 2:1. The synthesis reactions were analyzed with analytical RP-HPLC (annotation of the peaks was done using ESI-MS analysis of peak fractions). *p*-BABA-ATP always gave a double peak in the RP-HPLC analysis (Figure 3A), which is most likely due to the coupling of the biotinylated linker to the 2' and 3'-OH group of the ribose ring of ATP. With an eightfold excess of *p*-BABA hardly any *p*-BABA-ATP was formed (data not shown). The formation of *p*-BABA-ATP increased with increasing ATP up to equimolar concentrations, indicating a high efficiency for the activation process of the biotinylated linker. The reaction time was varied from 1 to 25 hours (Figure 3B) with the highest yield after 25 hours. Interestingly, the *p*-BABA-ATP isomer of the first peak was formed first and the other isomer only after longer conversion times (>1h).

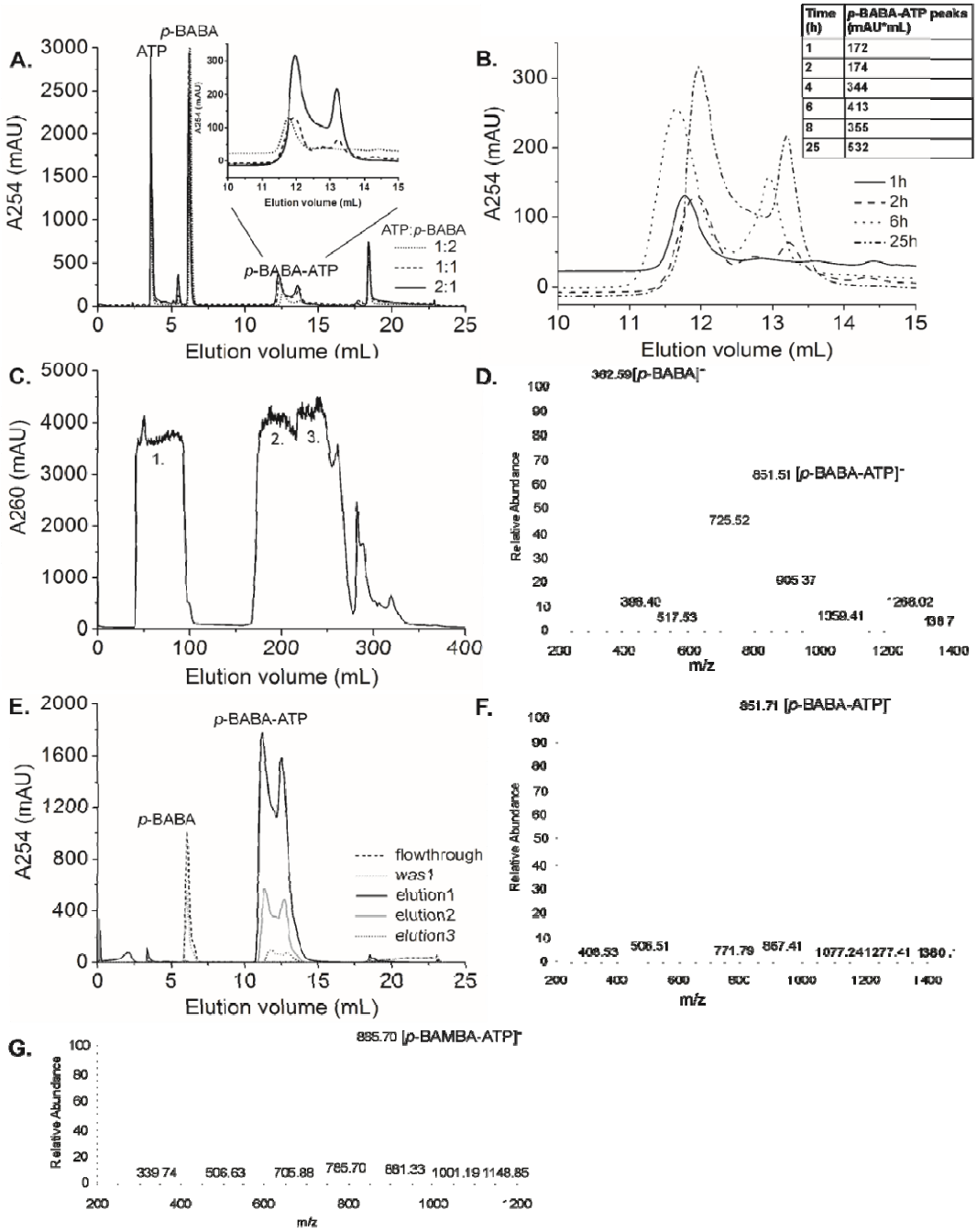


Figure 3. Characterization of the reaction products from the syntheses of the two ATP analogues. Annotation of the peaks in the HPLC measurements was based on standard runs with ATP and *p*-BABA and/or ESI-MS analysis of the peak fractions. Panel A: aRP-HPLC analysis of the *p*-BABA-ATP synthesis at different ratios of the starting materials, that are, ATP and *p*-BABA (inset: enlargement of the *p*-BABA-ATP double peaks); 20x diluted samples of the reaction mixtures are shown. Panel B: aRP-HPLC analysis of the *p*-BABA-ATP synthesis as a function of time; 100x diluted samples are shown (inset table: the summed values of the integral of both *p*-BABA-ATP peaks). Panel C: spectrum of a pRP-HPLC purification of the crude *p*-BABA-ATP reaction mixture. Peak 1 contains ATP, peak 2 contains *p*-BABA and peak 3 contains *p*-BABA-ATP, all based on ESI-MS analysis. Panel D: ESI-MS (negative mode) analysis of the pooled fractions of pRP-HPLC purified *p*-BABA-ATP (fractions from peak 3, panel C). Panel E: aRP-HPLC spectra from a monoQ purification of the pRP-HPLC purified *p*-BABA-ATP reaction mixture. Panel F: ESI-MS (negative mode) analysis of the pooled elution fractions of the *p*-BABA-ATP after monoQ purification (elution fractions from panel F after multiple rounds of freeze drying to remove the TEAA buffer). Panel G: ESI-MS (negative mode) analysis of monoQ-purified *p*-BAMBA-ATP.

Two purification steps were performed to obtain the genuine ATP analogues. In the preparative RP-HPLC (pRP-HPLC), the peaks of *p*-BABA and *p*-BABA-ATP were not completely separated (Figure 3C) and a mixture of *p*-BABA and *p*-BABA-ATP was detected by ESI-MS in the pooled fractions after the first separation (Figure 3D), but ATP was no longer present. The pRP-HPLC purified *p*-BABA/*p*-BABA-ATP mixture was fractionated further on a strong anion exchange resin (monoQ sepharose). *p*-BABA did not bind to the column material (Figure 3E). *p*-BABA-ATP eluted at 1M TEAA pH 7 and the ESI-MS analysis of the pooled elution fractions (elution fractions 1-3) showed a clean signal for the ATP analog (Figure 3F).

In general, step-elution was used to obtain the pure ATP analog in relative small fractions; with a linear gradient it was possible to separate the two different isomers (data not shown). The second ATP analog *p*-BAMBA-ATP behaved similarly both on the RP-HPLC and on the monoQ sepharose column, and the analysis of the purified fractions by ESI-MS confirmed the identity of *p*-BAMBA-ATP (Figure 3G).

Structural properties

The ESI-MS spectra of the product proved that the coupling of *p*-BABA to ATP was successful but did not establish the position of the coupling, that is, the 2'-OH or 3'-OH position of the ribose ring, the γ-phosphate or the amine-group of the adenine ring). COSY-NMR was tried to facilitate the annotation of the ^{1}H-NMR spectrum, but the spectra were too complex. Measurements with ^{13}C-NMR (^{13}C-NMR and HSQC) were limited by the low solubility of the ATP analog. The spectra did not improve enough at higher temperatures or longer measurement times (up to 1 week). In order to reduce the complexity of the NMR spectra, the synthesis was setup with 2-deoxy-ATP (dATP) instead of ATP. Addition of the dATP did affect the reaction significantly, for instance a precipitate was already formed during the synthesis. Chromatography of the soluble part of the synthesis mixture yielded an elution pattern similar to *p*-BABA-ATP. *p*-BABA-dATP was present in the peak fractions, but both starting materials (*p*-BABA and dATP) were present as well. Since, the pRP-HPLC purified was designed to separate the ATP (and dATP elutes even earlier than ATP, data not shown) from the *p*-BABA and *p*-BABA-dATP, the data imply that the dATP must originate from the degradation of *p*-BABA-dATP.

A variety of other methods were used to demonstrate the coupling of the biotinylated linker to the hydroxyl groups of the ribose ring of ATP. Measurements of *p*-BABA-ATP and ATP with ^{31}P-NMR showed equal spectra (data not shown), which excludes coupling of the *p*-BABA to the γ-phosphate of ATP. Partial hydrolysis and detection by ESI-MS (and MSMS) of *p*-BABA-ATP was used to distinguish between coupling of the biotinylated linker to the amine-group of the adenine ring or the hydroxyl-groups of the ribose ring. Conditions were optimized to retain the amide bonds (which could be checked easily because biotin is also coupled via an amide bond to the linker) and to hydrolyze the ester bond. The ester bonds can be hydrolyzed both via acid

or base catalyzed reactions. Incubation with 1M HCl at 100 °C for 1h did hydrolyze *p*-BABA-ATP but also hydrolyzed the amide bond between biotin and the linker (data not shown). Further experiments were performed with a base-catalysed reaction. The hydrolysis with base generates a transesterification reaction with the solvent (methanol). The ESI-MS spectrum of the ATP analog after transesterification with the base sodium-carbonate in methanol did show hydrolysis of *p*-BABA-ATP (Figure 4A&B) and the methylester of *p*-BABA could be detected. Reaction of the amide bonds under these conditions is highly unlikely, which supports that *p*-BABA is attached to the 2'/3'-OH groups via an ester linkage. The peak of the methylester of *p*-BABA was analyzed further by ESI-MSMS to prove the transesterification event (Figure 4C), which clearly showed the fragments of the biotinylated linker. The addition of formic acid, needed for charging of the molecules for the MS measurements did increase the background signal in the ESI-MS spectrum (Figure 4B), showing a regular pattern of peaks 68 Da apart, which is most like formed by sodium adducts of formic acid. Beneath the sodium-formic acid salts peaks in Figure 4B several smaller peaks were detected, which correspond to ATP (Na-adduct), not-hydrolyzed ATP analog (Na-adduct) and *p*-BABA-ADP, indicating that the phosphates are partly hydrolyzed as well. In summary, the chemical analysis of *p*-BABA-ATP indicates that the biotinylated linker is indeed coupled to the ribose moiety of ATP (as presented Figure 2B).

Spectral properties

UV spectra of the different compounds were recorded to determine the absorption maximum and extinction coefficient of *p*-BABA-ATP and *p*-BAMBA-ATP. The absorption maxima of the two ATP analogues were different compared to the starting materials (Figure 5A). *p*-BABA-ATP (264 nm) had a red-shifted spectrum compared to ATP (258 nm) and *p*-BABA (260 nm), while the maximum of *p*-BAMBA-ATP (245 nm) was in between the optimum of ATP and *p*-BAMBA (230 nm). The molar extinction coefficient was not altered upon modification of ATP and under the experimental conditions ATP, *p*-BABA-ATP and *p*-BAMBA-ATP all had molar extinction coefficients of around 15,000 M^{-1} cm^{-1} at their respective absorption maxima.

Nucleotide-binding properties

A competition assay was used to determine the binding properties of *p*-BABA-ATP and *p*-BAMBA-ATP compared to ATP. The purified ABC transporter OpuA from *Lactococcus lactis* [27] was used as a model protein to assess the relative binding affinities. The nucleotide binding affinity of OpuA for ATP is in the millimolar range and cannot be determined directly. The fluorescent ATP analog trinitrophenyl-adenosine triphosphate (TNPATP) binds to OpuA with micromolar affinity and can be chased with ATP or other analogues to assess their affinity [28] (Figure 5B). Both *p*-BABA-ATP and *p*-BAMBA-ATP were at least 10-fold more efficient in chasing TNP-ATP from the binding pocket of OpuA than ATP; *p*-BABA-ATP displayed the highest affinity. These measurements show that

the linker does improve the binding affinity of the nucleotide analogues for the ABC transporter OpuA.

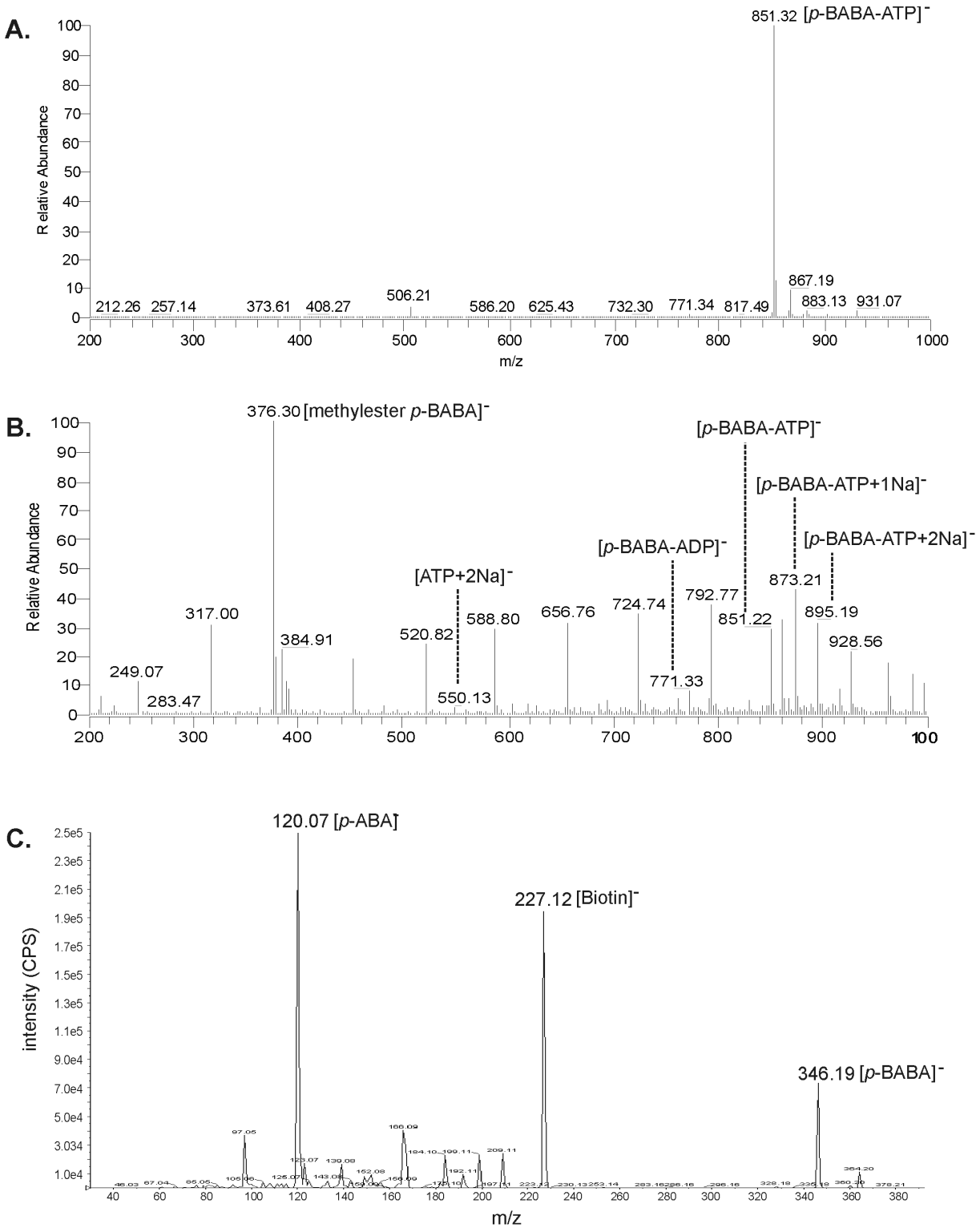


Figure 4. Analysis of the partial hydrolysis of *p*-BABA-ATP. A. ESI-MS (negative mode) of *p*-BABA-ATP dissolved in methanol. B. ESI-MS (negative mode) of *p*-BABA-ATP hydrolyzed in sodium-carbonate after 1h incubation at 40 °C. C. ESI-MSMS (positive mode) peak analysis of the *p*-BABA methylester.

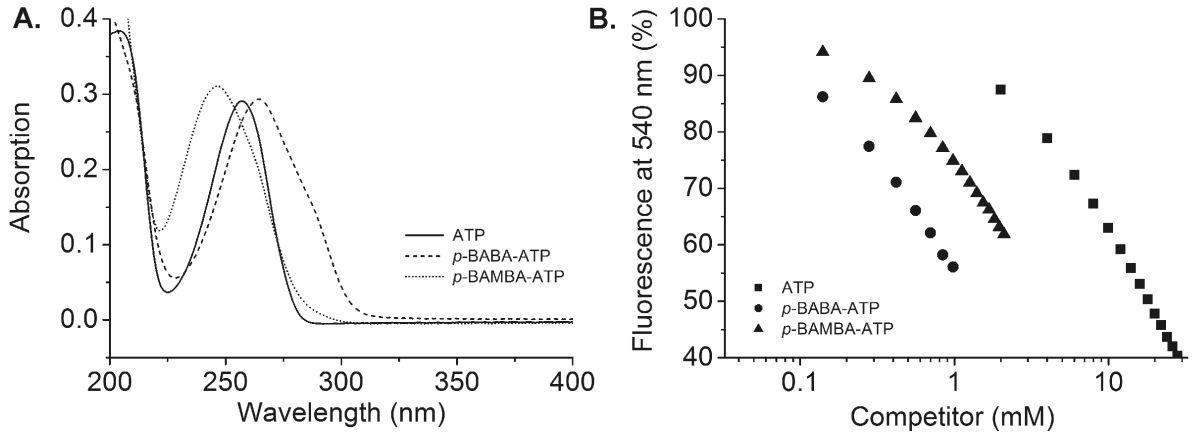


Figure 5. Charactization of the ATP analogues. Panel A: spectral properties of ATP, *p*-BABA-ATP and *p*-BAMBA-ATP (20 µM). The absorption reached an optimum at 245 nm for *p*-BAMBA-ATP, 258 nm for ATP and 264 nm for *p*-BABA-ATP in 50 mM KP_i pH 7. Panel B: nucleotide-binding properties of the ATP analogues. Binding of ATP or the ATP analogues was measured by chasing TNP-ATP bound to the purified ABC transporter OpuA.

Conclusions

Two ATP analogues with biotinylated linkers coupled to the alcohol-groups of the ribose ring of ATP were synthesized and purified successfully, namely biotinylamido benzoic acid-ATP (*p*-BABA-ATP) and biotinylaminomethyl benzoic acid-ATP (*p*-BAMBA-ATP). The coupling of biotin to the linker was demonstrated by ^{1}H- and ^{13}C- NMR, and the correct coupling of the biotinylated linker to the ATP was deduced from ESI-MS/MSMS analyses of partial digests of the ATP analogues and ^{31}P-NMR. Ligand-binding experiments showed that the two ATP analogues bind to the ABC transporter OpuA with at least one order of magnitude higher affinity than genuine ATP.

Material and Methods

Synthesis of biotinylated linkers

The synthesis of biotinylated linkers is based on a description of biotinylated amino acid precursors by Skander et al. [24]. Briefly, 2.05 mmol biotin was disssolved in 40 mL DMF with 2.69 mmol tributylamine (0.64 mL). 2.47 mmol isobutyl chloroformate (0.32 mL) was added to the solution to activate the biotin, and the mixture was incubated at room temperature for about 15 minutes after which it was cooled on ice. Meanwhile, 4.1 mmol linker was dissolved in 40 mL DMF and cooled on ice. After cooling, the activated biotin mixture was slowly added to the linker dissolved in DMF, and the mixture was incubated for two hours at 5 °C. The solvent was removed under vacuum and the crude precipitate was recrystallized from 3:1 ethanol/H_2O, and the obtained crystals were washed extensively with water and ethanol. The reaction was performed with two different linkers: *p*-aminobenzoic acid (*p*-ABA) and *p*-aminomethylbenzoic acid (*p*-AMBA). Both biotinylated linkers recrystallized as a white solid, but the recrystallized solid of the biotinylated *p*-aminomethylbenzoic acid was more hygroscopic compared to

the solid of the biotinylated *p*-aminobenzoic acid. The yield of pure biotinylated linker is about 40 % and purity of the products was analyzed by NMR:
p-biotinyl-amidobenzoic acid (p-BABA): ^{1}H-NMR (300 MHz, CD_3SOCD_3): δ = 1.34-1.62 (m, 6H), 2.33 (t, *J* = 7.3 Hz, 2H), 2.56 (d, *J* = 12.5, 1H), 2.80 (dd, *J* = 12.5, 5.1 Hz, 1H), 3.07-3.13 (m, 1H), 4.12 (t, *J* = 5.1 Hz, 1H), 4.29 (t, *J* = 6.2 Hz, 1H), 6.39 (d, *J* = 23.4 Hz, 2H), 7.68 (d, *J* = 8.8 Hz, 2H), 7.85 (d, *J* = 8.8, 2H). ^{13}C-NMR (200 MHz, CD_3SOCD_3): δ = 25.1, 28.3, 36.5, 55.5, 59.4, 61.3, 118.5, 125.1, 130.5, 143.4, 163.0, 167.2, 172.1.
p-biotinyl-amidobenzoic acid (p-BAMBA): ^{1}H-NMR (300 MHz, CD_3SOCD_3): δ 1.31-1.57 (m, 6H), 2.14 (t, *J* = 7.1 Hz, 2H), 2.56 (d, *J* = 12.5 Hz, 1H), 2.81 (dd, *J* = 12.5, 5.1 1H), 3.06-3.09 (m, 1H), 4.11 (t, *J* = 5.1 Hz, 1H), 4.27-4.31 (m, 3H), 6.39 (d, *J* = 21.2 Hz, 2H), 7.33 (d, *J* =8.4 Hz, 2H), 7.87 (d, *J* = 8.4 Hz, 2H), 8.41 (t, *J* = 5.5 Hz, 1H). The NMR spectra can be found in Supplement III.

Synthesis biotinyl-linker-ATP
The synthesis of the biotinyl-linker-ATP complex is based on the activation of an acid with carbonyldiimidazole [26] and subsequent coupling to an alcohol group. This method to couple an aromatic moiety to ATP is based on the synthesis of 3'-O-(4'-Benzoyl)benzoyl-ATP as described by Williams et al. [25]. Glassware was dried at 160 °C prior to use. 0.48 mmol biotinylated linker was dissolved in 3 mL dry DMF together with 1.18 mmol carbonyldiimidazole (CDI), keeping the setup with the dried glassware under a flow of nitrogen. The reaction was monitored by the evaporation of CO_2-gas, which stopped after 1h after addition of CDI. Then, 0.48 mmol of ATP in 12 mL H_2O was added to the mixture and the reaction was incubated for at least 6 hours under constant stirring at room temperature.

The crude reaction mixture was purified in batches of 5 mL by RP-HPLC, using a preparative C18 column (prepared from Vydac218TPB_1015RPC column material; column dimensions 25/132) connected to a Äkta-purifier system. The column was run at a flow rate of 3 mL/min, using buffer A: 10 mM triethylammoniumacetate (TEAA) pH 7 and buffer B: acetonitrile plus 10 mM TEAA pH 7, and the following gradient: 35 min 5% buffer B, 75 min gradient 5-35% buffer B, 15 min 90% buffer B, 15 min 0% buffer B. 15 mL fractions were collected and analyzed by ESI-MS to determine the fractions containing the ATP analog. Peak fractions of multiple runs were pooled and freeze-dried and the material was then resuspended in 200 mM TEAA pH 7 and loaded onto a 1 mL monoQ sepharose column, pre-equilibrated with 20 column volumes 200 mM TEAA pH 7. The column was washed with 2x10 column volumes 200 mM TEAA pH 7 and eluted with 4x5 column volumes 1M TEAA pH 7.0. Multiple rounds of freeze-drying (the solids were resuspended in milliQ) were done to remove the (volatile) TEAA pH 7 completely. The yield of pure product was about 20% and the final products were analyzed with ESI-MS:

p-biotinyl-amidobenzoic acid-ATP (p-BABA-ATP): (ESI-MS), calculated mass $C_{27}H_{35}N_8O_{16}P_3S$ 852,61, found *m/z* $[M]^-$ 851.71 (50 % ACN + 0.1% formic acid)/ 851.32 (methanol).
p-biotinyl-amidobenzoic acid-ATP (p-BAMBA-ATP*)*: (ESI-MS), calculated mass $C_{28}H_{37}N_8O_{16}P_3S$ 866.64, found *m/z* $[M]^-$ 865.60 (50 % ACN + 0.1% formic acid).

Partial hydrolysis of p-BABA-ATP
A few mg of *p*-BABA-ATP was dissolved in 250 µL methanol. Sodium carbonate was added until the total volume was filled about halfway with powder, to ensure saturating conditions. The mixture was incubated for 1 hour at 40 °C, under mixing at 800 rpm. The mixture was shortly spun down and the supernatant was mixed 1:1 with 50 % ACN plus 0.1 % formic acid. Formic acid was added for the ESI-MS analysis. For ESI-MS/MSMS analysis, the samples were diluted 10x in 50% ACN plus 0.1 % formic acid.

Analytical methods
Nuclear magnetic resonance spectra were recorded with a 200/300 MHz (^{1}H-NMR) or 50/75 MHz (^{13}C-NMR) spectrometer. Chemical shifts are reported in delta (δ) units, parts per million (ppm), and as an average shift for split peaks. Coupling constants (*J*) are reported in Hz. ESI-MS spectra were measured with a LCQ fleet mass spectrometer (Thermo scientific) via direct injection at a flow rate of 10 µL/min. Settings of the spectrometer were tuned with a standard solution of ATP both in positive and negative ion mode. Samples were diluted in 50% acetonitrile plus 0.1% formic acid prior to the measurements. ESI-MS/MSMS mass spectra were recorded on an API3000 triple quadrupole mass spectrometer equipped with a TurboIonSpray source (MDS-Sciex, Concord, Ontario, Canada) in positive ion mode with nitrogen as collision gas, at a collision energy of 40 eV.

Analytical RP-HPLC was performed using a polystyrene/divenyl benzene column (Source^{TM} 5RPC ST 4.6/150) connected to a Äkta-purifier system. Two buffers were used for the separation: Buffer A: 100 mM TEAA pH 7.0 and Buffer B: acetonitrile plus 100 mM TEAA pH 7.0. The settings for the separations were: 200 µL loop, flow 1 mL/min, gradient: 5 min 17% buffer B, 10 min gradient 17-20% buffer B, 5 min 90% buffer B, 5 min 0%B.

Absorption spectra were measured on a Cary100 Bio UV-visible spectrophotometer (Varian) with samples of 400 µL in a quartz cuvette.

TNP-ATP binding
0.7 µM purified OpuA was pre-incubated for 10 minutes with 2 µM TNP-ATP in 50 mM KP_i pH 7 plus 20% glycerol, 200 mM KCl and 0.04% DDM. Fluorescence was measured in a total volume of 800 µL under constant stirring at 25 °C on a Fluorolog-3 spectrophotometer (Jobin Yvon, settings: excitation wavelength of 409 nm (slit 1 nm), emission wavelength of 540 nm (slit 5 nm), and data points were averaged from measurements of 20 sec). Nucleotides to chase the TNP-ATP were added via a Hamilton

syringe pump (Harvard apparatus) in steps of 4.5 µL of 2 mM for ATP and 0.14 mM for the ATP analogues (using stock solutions of 357 mM ATP and 25 mM ATP for the analogues, both dissolved in 50 mM KP_i pH 7) and a mixing time of 15 sec. Titration of protein solution with 50 mM KP_i pH 7 were performed to correct for dilution and bleaching during the measurements.

Acknowledgements

We would like to thank Hjalmar Permentier for his help with the ESI-MS/MSMS measurements, and Ruud Scheek and Renee Otten for their help with the NMR measurements. This work was supported by grants from The Netherlands Organisation for Scientific research (NWO, Chemical Sciences Top Subsidy to BP; grant number 700-56-302 and the Netherlands Proteomics Centre (NPC).

References

[1] Laurie,A.T. & Jackson,R.M. (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics.*, **21,** 1908-1916.

[2] Bagshaw,C.R. (2001) ATP analogues at a glance. *J. Cell Sci*, **114,** 459-460.

[3] Cremo,C.R. (2003) Fluorescent nucleotides: Synthesis and characterization. *Biophotonics, Pt A*, **360,** 128-177.

[4] Jameson,D.M. & Eccleston,J.F. (1997) Fluorescent nucleotide analogs: Synthesis and applications. *Fluor. Spectr*, **278,** 363-390.

[5] Yount,R.G. (1975) ATP analogs. *Adv. Enzymol Relat Areas Mol Biol*, **43,** 1-56.

[6] Leonard,N.J., Scopes,D.I., VanDerLijn,P., & Barrio,J.R. (1978) Dimensional probes of the enzyme binding sites of adenine nucleotides. Biological effects of widening the adenine ring by 2.4 A. *Biochemistry*, **17,** 3677-3685.

[7] Van der Hijden,H.T., Kramer-Schmitt,S., Grell,E., & de Pont,J.J. (1990) The basal Mg2(+)-dependent ATPase activity is not part of the (H(+)+K+)-transporting ATPase reaction cycle. *Biochem. J*, **267,** 565-572.

[8] Eckstein,F. & Goody,R.S. (1976) Synthesis and properties of diastereoisomers of adenosine 5'-(O-1-thiotriphosphate) and adenosine 5'-(O-2-thiotriphosphate). *Biochemistry*, **15,** 1685-1691.

[9] Colman,R.F. (1983) Affinity labeling of purine nucleotide sites in proteins. *Annu. Rev. Biochem.*, **52,** 67-91.

[10] Czarnecki,J.J. (1984) Tautomerism of 2-azidoadenine nucleotides. Effects on enzyme kinetics and photoaffinity labeling. *Biochim. Biophys Acta*, **800,** 41-51.

[11] Bayley,H. & Knowles,J.R. (1977) Photoaffinity labeling. *Methods Enzymol*, **46,** 69-114.

[12] Pal,P.K., Wechter,W.J., & Colman,R.F. (1975) Affinity labeling of the inhibitory DPNH site of bovine liver glutamate dehydrogenase by 5'-fluorosulfonylbenzoyl adenosine. *J Biol Chem*, **250,** 8140-8147.

[13] Pal,P.K., Wechter,W.J., & Colman,R.F. (1975) Affinity labeling of a regulatory site of bovine liver glutamate dehydrogenase. *Biochemistry*, **14,** 707-715.

[14] Pal,P.K. & Coleman,P.S. (1990) Detecting precatalytic conformational changes in F1-ATPase with 4-benzoyl(benzoyl)-1-amidofluorescein, a novel fluorescent nucleotide site-specific photoaffinity label. *J Biol Chem*, **265,** 14996-15002.

[15] Easterbrook-Smith,S.B., Wallace,J.C., & Keech,D.B. (1976) Pyruvate carboxylase: affinity labelling of the magnesium adenosine triphosphate binding site. *Eur. J Biochem*, **62,** 125-130.

[16] Green,N.M. (1963) Avidin .1. Use of [14C]Biotin for Kinetic Studies and for Assay. *Biochem. J*, **89,** 585-&.

[17] Takenaka,H., Ikehara,M., & Tonomura,Y. (1978) Interaction between actomyosin and 8-substituted ATP analogs. *Proc. Natl. Acad. Sci U. S. A*, **75,** 4229-4233.

[18] Onodera,M. & Yagi,K. (1971) Synthesis of 2-(dansylamino)ethyl triphosphate and its properties as a fluorescent substrate of heavy meromyosin-ATPase. *Biochim. Biophys Acta*, **253,** 254-265.

[19] Schmidt,J.A. & Colman,R.F. (1984) Identification of the lysine and tyrosine peptides labeled by 5'-p-fluorosulfonylbenzoyladenosine in the NADH inhibitory site of glutamate dehydrogenase. *J Biol Chem*, **259,** 14515-14519.

[20] Schafer,H.J., Coskun,U., Eger,O., Godovac-Zimmermann,J., Wieczorek,H., Kagawa,Y., & Gruber,G. (2001) 8-N-3-3 '-biotinyl-ATP, a novel monofunctional reagent: Differences in the F-1- and V-1-ATPases by means of the ATP analogue (vol 286, pg 1218, 2001). *Biochem. Biophys. Res. Commun*, **288,** 490-492.

[21] Hiratsuka.T & Uchida,K. (1973) Preparation and Properties of 2'(Or 3')-O-(2,4,6-Trinitrophenyl) Adenosine 5'-Triphosphate, An Analog of Adenosine-Triphosphate. *Biochim. Biophys. Acta*, **320,** 635-647.

[22] Smith,P.C., Karpowich,N., Millen,L., Moody,J.E., Rosen,J., Thomas,P.J., & Hunt,J.F. (2002) ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell*, **10,** 139-149.

[23] Horn,C., Bremer,E., & Schmitt,L. (2003) Nucleotide dependent monomer/dimer equilibrium of OpuAA, the nucleotide-binding protein of the osmotically regulated ABC transporter OpuA from Bacillus subtilis. *J. Mol. Biol*, **334,** 403-419.

[24] Skander,M., Humbert,N., Collot,J., Gradinaru,J., Klein,G., Loosli,A., Sauser,J., Zocchi,A., Gilardoni,F., & Ward,T.R. (2004) Artificial metalloenzymes: (Strept)avidin as host for enantioselective hydrogenation by achiral biotinylated rhodium-diphosphine complexes. *J. Am. Chem. Soc*, **126,** 14411-14418.

[25] Williams,N. & Coleman,P.S. (1982) Exploring the Adenine-Nucleotide Binding-Sites on Mitochondrial F1-Atpase with A New Photoaffinity Probe, 3'-O-(4-Benzoyl)Benzoyl Adenosine 5'-Triphosphate. *J. Biol. Chem*, **257,** 2834-2841.

[26] Gottikh,B.P., Krayevsky,A.A., Tarussova,N.B., Purygin,P.P., & Tsilevich,T.L. (1970) The general synthetic route to amino acid esters of nucleotides and nucleoside-

5'-triphosphates and some properties of these compounds. *Tetrahedron*, **26,** 4419-4433.

[27] Patzlaff,J.S., van der Heide,T., & Poolman,B. (2003) The ATP/Substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem*, **278,** 29546-29551.

[28] Wolters,J.C., Berntsson,R.P., Gul,N., Karasawa,A., Thunnissen,A.M., Slotboom,D.J., & Poolman,B. (2010) Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLoS. One*, **5,** e10361.

Chapter 3 Enrichment of adenine nucleotide-binding proteins using the ATP-based nucleotide analogues

Justina C. Wolters, Gerard Roelfes and Bert Poolman

Abstract

The ability of the two ATP analogues *p*-BABA-ATP and *p*-BAMBA-ATP, described in Chapter 2 to be used for specific enrichment of adenine-nucleotide binding proteins is described in this chapter. Several overproduced ABC transporters could be enriched specifically via the biotin moiety on the nucleotide analogues. Though in complex proteins mixtures, the affinity to the ATP analogues seems to be too low to discriminate between specific binding of adenine-nucleotide binding proteins and aspecifically bound proteins.

Introduction

Activity-based proteomics

In activity-based protein profiling (proteomics), protein classes are enriched via a reactive probe that requires the protein or protein domain to be functional. The probe (bait) contains a reactive group and a reporter tag, usually connected via a linker moiety. The binding moiety is designed to react with a specific class of (low abundant) proteins and the reporter tag of the probe enables their detection and/or enrichment (Figure 1). The use of different binding moieties and reactive groups can help in the analysis of the active site of the proteins. The probes can be used to screen for (unlabeled) inhibitors via competitive binding studies, with the clear advantage that experiments can be performed in complex mixtures without the requirement for prior enrichment steps (e.g. as used for the identification of drug targets [1]). Numerous reviews about activity-based proteomics and their applications have been published in recent years (2006-2008) [2-8]. Here, an overview of the different strategies of activity-based proteomics is presented.

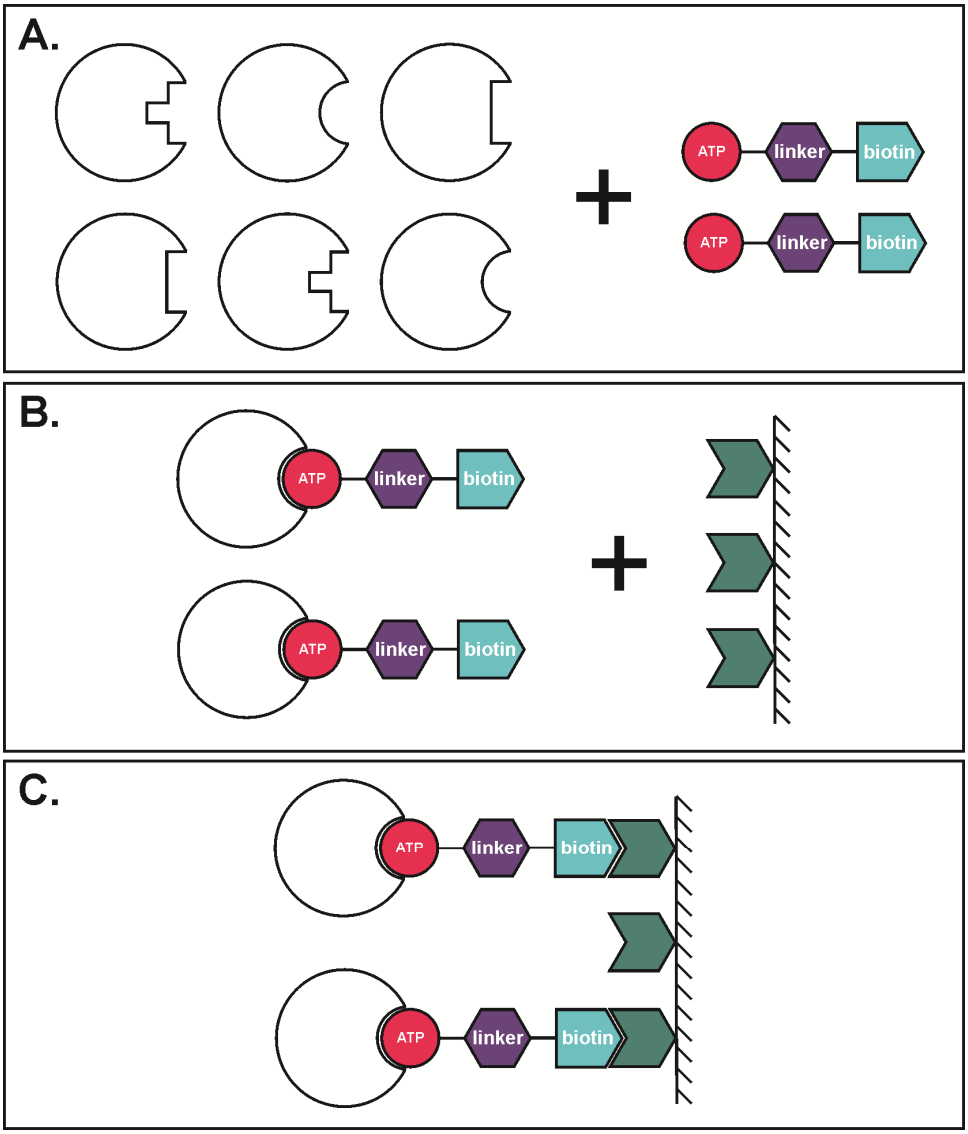


Figure 1. Schematic representation of activity based profiling using the ATP analog as probe. The ATP (reactive group) moiety of the probe binds specifically to nucleotide binding proteins within a complex sample (panel A). The biotin (reporter group) moiety of the probe is used for enrichment of the proteins bound to the ATP analog to Streptactin column material (panel B). The proteins that are enriched via this method can be analyzed after elution from the Streptactin column (panel C).

Detection methods

The presence of reporter tags (fluorescent, affinity, antigen or isotope-labeled tags) enables screening of labeled proteins or inhibitor/competitor studies. Identification of proteins bound to the probes is usually done via mass spectrometry (MS). Two major methods have been applied to reduce the sample complexity either via gel-based or liquid chromatography (LC)-based separations in one or two dimensions. Gel-based separations techniques have the advantage that large numbers of proteomes can be screened fast and in parallel. LC-MS methods have a superior resolution over gel-based separation techniques, but experiments may take longer. Jessani et al. [9] used a combination of both approaches, using fast initial identification of targets on gels and in depth analysis of representative members by LC-MS.

The advantage of a covalently-linked probe is that proteins remain labeled upon unfolding. The unfolding can improve the accessibility of the probe for column binding. The enrichment can be done on protein level (yielding all active proteins) or after digestion of proteins to peptides (yielding labeled peptides only, so-called active-site peptide profiling). The disadvantage of the active-site peptide profiling is that identification of the proteins is limited by the identification of the labeled peptide only. This limitation was overcome by Speers and co-workers who combined two techniques in a method they called tandem orthogonal proteolysis activity-based protein profiling (TOP-ABPP) [10]. Labeled proteins were enriched on a column and digested with trypsin, while bound to the column to collect the peptides for protein identification. The labeled peptides were subsequently eluted from the column and identified, using a cleavable Tobacco etch virus protease (TEV) site introduced in the probe.

Protein profiling is routinely done in cell homogenates rather than *in vivo*, because the size of the probes limits their uptake and distribution in the cells. This limitation can be overcome by splitting of the label into two parts: a crosslinking group and a reporter tag (fluorescent probe or biotin tag). Both parts of the probe contain an extra reactive group to attach the two halves, using so-called click chemistry (click chemistry in general, see [11]; click chemistry implemented in activity-based probes, see [12,13]). In a click chemistry reaction, the two parts of the probes contain an azide and alkyne group, respectively, which can form a stable triazole product in a copper(I)-catalyzed cyclo-addition reaction. In vivo labeling has been demonstrated, but the levels of background signals vary depending on the position of the azide and alkyne labels on the probe parts [12,13].

Activity-based probes

Activity-based protein profiling experiments have been done for various protein classes, like hydrolases, proteases, phosphatases, kinases and glycosidases. These protein classes have been analyzed, because they are potential targets for the development of new drugs. Different types of activity- or mechanism-based probes are discriminated, like activity- or affinity based probes [6] or directed and non-directed probes [5]. Directed probes are targeted to the active site using an inhibitor or ligand. These probes can be either irreversibly or reversibly coupled in the binding site. The irreversible probes are modified through enzymatic activity to form the covalent cross-link, hence the name activity-based protein profiling. The binding moiety of the reversible probes acts as a scaffold for interaction with specific protein classes, but a reactive group needs to be added to crosslink the probes to the proteins, which is usually achieved via (photo)-chemical crosslinking. The non-directed probes contain a more general reactive binding group, like sulfonate esters or chloroacetamides, which can be used to screen multiple protein classes at the same time. Probes that have been prepared and validated for analysis of different protein classes are discussed below.

Directed irreversible probes

Most directed irreversible probes are designed to attack nucleophilic residues in the catalytic binding pocket and cross-linking is achieved through enzymatic activity. Classifications can be made on the basis of the reactive groups used for the covalent linkage, like quinone methide, ketones, phosphates/phosphonates, and sulfones/sulfonates. Probes containing a quinone methide or quinol-imine have been used for the screening of the protein tyrosine phosphatases [14,15] (with dansyl or Cy3 fluorescent groups or biotin attachment), glycosidases [16] (with a biotin moiety attached) and proteases (using the fluorescent Cy3-moiety; protease subclasses were differentiated via the use of various inactivators) [17]. Glycosidases have been also studied with probes that target the catalytic nucleophilic carboxylate [18,19], containing a dansyl-group or a biotin, coupled either via a cleavable or non-cleavable linker. Electrophilic fluorophosphonates react with the serine in the binding pocket of serine hydrolases [20,21], with biotin attached with or without a PEG-linker, or serine proteases [22,23],with a fluorescein-moiety attached via click chemistry or coupling to a biotin-group. Other phosphonate-based probes[1] like *p*-nitrophenyl phosphonates [24] and bromobenzylphosphonate [25,26] have been used to differentiate between lipases and esterases, and to study protein tyrosine phosphatases respectively (with biotin or rhodamine groups attached). Protein tyrosine phosphatases have been also studied using aryl vinyl sulfonate and sulfone, taking advantage of click chemistry to attach biotin-based probes [27]. This class of probes has better properties for *in vivo* studies than the other ones, since the quinone methides are not specific for protein tyrosine phosphatases and the phosphonates have low membrane permeability [27]. Vinylsulfone-based probes have been used for the detection of cysteine proteases [28-30] and the proteasome [31].

Probes designed for targeting cysteine proteases, contain an epoxide-based reactive functionality [32] (with a biotin-group attached), and these types of probe have been also modified with stable isotope labeling (ICAT) for quantitative studies [33]. Probes designed with ketone-based structures like fluoromethyl ketones [34] and (acyloxy)methyl ketones [35,36], and containing fluorescent groups, biotin and/or radio-isotopes for detection purposes, have been used to profile cysteine proteases.

Miscellaneous probes: β-lactam-derived probes have been prepared to characterize penicillin-binding proteins (PBPs), exploiting the electrophilic β-lactone ring to attack a nucleophilic residue in the active site [37,38], using click chemistry for detection purposes. Inhibitors like 2-ethynylnaphthalene have been used for screening of cytochrome P450 monooxygenases [39]. The ATP mimic 5'-p-fluorosulfonylbenzoyl

[1] Literature on the profiling of cysteine proteases by phosphonate-based probes gives contradictory information. In in the review by Schmidinger [6] , they are described as probes for profiling of cysteine proteins (though no references to support the claims were given), whereas the opposite is indicated in the review by Cravatt [3].

adenosine (5'-FSBA) has been used as irreversible inhibitor of dehydrogenases and kinases via alkylation with lysine amino-groups in the nucleotide-binding pocket [40].

Directed reversible probes

Reversible-probes containing a benzophenone moiety for covalent crosslinking have been used to profile proteases, including aspartyl protease activity [41]. Probes containing click chemistry handles for detection of the labeled proteins have been used for the characterization of histone deacetylase complexes (zinc-dependent metallohydrolases) [42,43] and metalloproteases in general [44].

An acyl-phosphate analog ATP probe, (+)-biotin-hex-acyl-ATP (BHacATP), containing a biotin moiety linked with a hex-acyl-linker to the phosphate [45], has been designed specifically for protein kinases with a conserved lysine in the binding pocket, i.e., near the phosphates of the nucleotide [46,47] (Figure 2A and 2C).

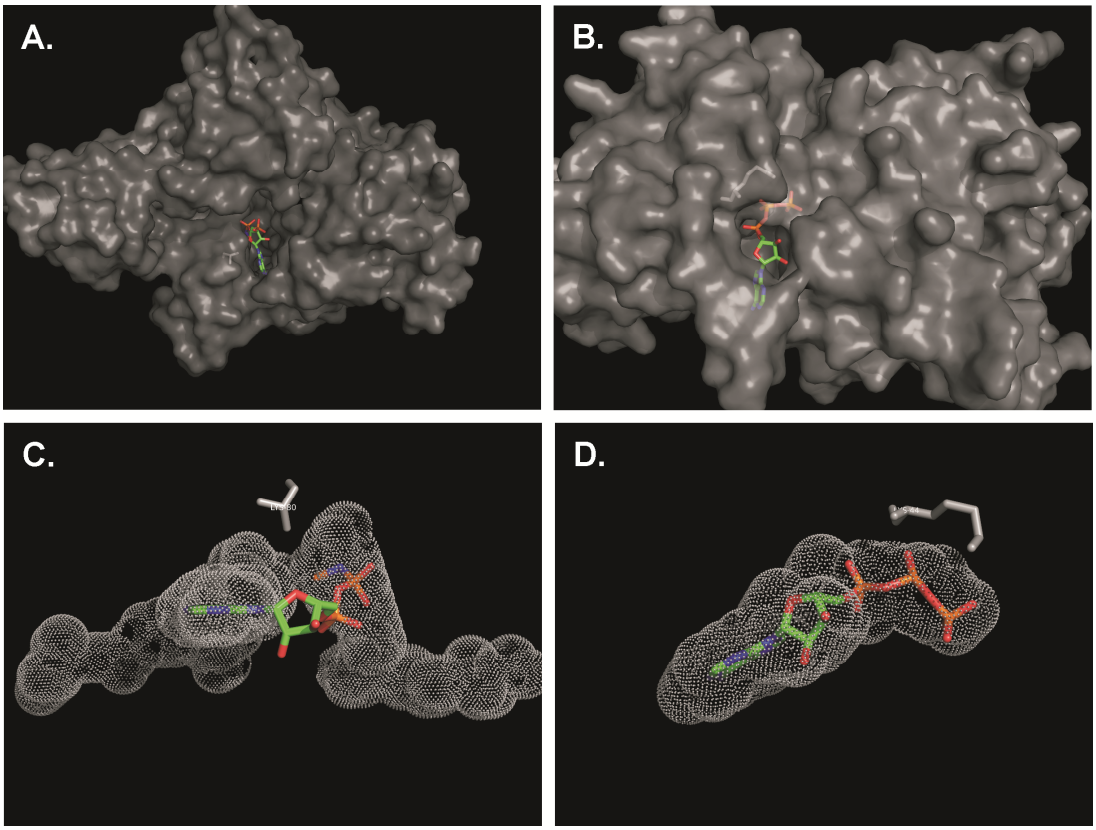


Figure 2. Crystal structures showing the position of conserved lysines in the nucleotide-binding pocket of a protein kinase (3KU2 panel A and C) and the ABC transporter MJ0796 (1L2T panel B and D). The position of the conserved lysine (K80) in the protein kinase is shown in the full structure (panel A) and relative to the modeled binding pocket (panel C), which shows space around the phosphate of the nucleotide (and thus the ability to accommodate phosphate-modified ATP analogues). The conserved lysine (K44) of the ABC transporter is part of the walker A motif present in all ABC transporters. The nucleotide-binding pocket of ABC transporters is too tight for modifications around the phosphate moiety.

The electrophilic acyl-phosphate bond will couple with the ε-amine from the lysine, connecting the biotinylated linker to the protein and freeing it from the nucleotide. Using this approach, proteome profiling of breast cancer cell lysates identified 136

proteins, mostly ATP phosphohydrolases, (protein) kinases and several heat shock proteins. Other proteins that have been identified belong to the families of NAD/FAD-dependent enzymes, GTPases and DNA/RNA-binding enzymes [45].

Non-directed probes

Adam *et al*. [48] has described the use of the first class of non-directed probes, that is, a library of probes with sulfonate esters as reactive group. These probes were targets for multiple dehydrogenases, a hydratase, a hydrolase, a phosphofructokinase and a glutathione S-transferase [48-50]. The phenylsulfonate probe has been also designed on the basis of click chemistry and used for *in vivo* detection [12,13]. Another non-directed probe is based on the reactivity of chloroacetamide, which couples to dipeptide-binding units [51], to obtain more structural diversity than with the sulfonate ester library. The probes were either coupled to rhodamine (w/wo biotin) for detection.

Experimental design for profiling of ABC (transport) proteins

The use of the ATP analog BHacATP did identify multiple adenine nucleotide-binding proteins in the profiling of a breast cancer proteome. On the basis of the data presented [45], ABC transporters were not detected. Figure 2B and 2D may explain why. Analysis of the the nucleotide-binding pockets of ABC transporters show the presence of a conserved lysine in the Walker A motif close to the phosphates of the nucleotide needed for the cross-linking. On closer inspection of the binding pocket, one observes that there is no space for the hexyl-linker with the biotin moiety, contrary to the situation in protein kinases. The probes designed and synthesized in Chapter 2 were designed for screening of adenine nucleotide-binding proteins and tuned for optimal detection of the family of ABC (transport) proteins. Detection of these proteins is important because ABC transporters form a large family of membrane proteins that are present in both microorganisms and eukaryotic species. They play an important role in a wide range of physiological functions, mostly involving translocation of compounds across the membrane [52]. In 2005, eighteen of the human ABC genes were associated to diseases, including cystic fibrosis (ABCC7), Tangier disease (ABCA1), immune deficiency (ABCB2/3) and many others (listed in review [53]; and the website of the human genome nomenclature committee (HGNC) also provides a lot information on ABC transporters in relation to diseases: http://nutrigene.4t.com/humanabc.htm).

We now present an approach to profile ABC proteins in complex cell lysates. Proteins in complex samples that bound to a Streptactin column material were identified via MALDI-MS/MSMS analysis, after trypsin digestion and reduction of sample complexity using reverse phase liquid chromatography. Relative quantification was achieved via labeling of the peptides with isobaric tags (iTRAQ). Peptides labeled with different isobaric tags have identical masses, but in the MS/MS analyses the reporter-groups are split off and used for relative quantification of all peptides. As a prove of principle, several ABC and non-ABC transporters were used for the characterization of the ATP analogues, prior to the screening of complex proteomes.

Results

The binding of *p*-BABA-ATP and *p*-BAMBA-ATP to the ABC transporters was demonstrated via ligand competition experiments, showing that both ATP analogues were more efficient in chasing the binding of the fluorescent ATP analog TNP-ATP than ATP (Chapter 2). Binding of *p*-BABA-ATP to the Streptactin column material was confirmed by analytical RP-HPLC analysis and binding of the proteins via the ATP analogues was confirmed by SDS-PAGE analysis, immunoblotting and MALDI-MS/MSMS analyses. Two adenine nucleotide-binding proteins were used for the characterization of the synthesized ATP analogues, namely the soluble nucleotide-binding domain (GlcV) of the glucose ABC transporter from *Sulfolobus solfataricus* (with a nanomolar affinity for ATP) and the ABC transporter OpuA (submillimolar to millimolar affinity for ATP).

Streptactin binding properties

Analytical RP-HPLC was used to monitor the ATP analogues and starting materials throughout the reaction scheme. *p*-BABA-ATP was incubated with streptactin column material and the flow through was collected and analyzed and compared with the starting material. Molar ratios of 10:1, 1:1 and 1:10 *p*-BABA-ATP: streptactin column materials were tested (the amount of streptactin column material was kept constant); the RP-HPLC spectra are shown in Figure 3. In the presence of equal or ten-fold excess of streptactin column material, about 75% of *p*-BABA-ATP was bound, indicating that it is difficult to occupy all binding sites of the column material. In the presence of a 10-fold excess of *p*-BABA-ATP, still about 30 % was bound, which is higher than the expected 10% and may reflect some aspecific binding.

Next, the effect of an excess of ATP analog upon binding of the ABC transporter OpuA and the nucleotide-independent RibU (used as a 'negative' control) to the streptactin was tested. Membrane vesicles containing these proteins were incubated with 50 or 500 nmol *p*-BABA-ATP together with Streptactin column material (with a binding capacity of the column material of 30 nmol). The amount of protein bound was analyzed by immunoblotting against the His-tag (Figure 4A, Table S1.1). Quantification of the signals with the AIDA software showed about equal binding of OpuA at a concentration of 50 or 500 nmol *p*-BABA-ATP. Subsequent experiments were mainly conducted at equimolar concentrations of column material (binding sites) and ATP analog. The different experiments are summarized in Table S1 (experiment numbers are listed together with the table number) and representative data are shown in the Figures 4-7.

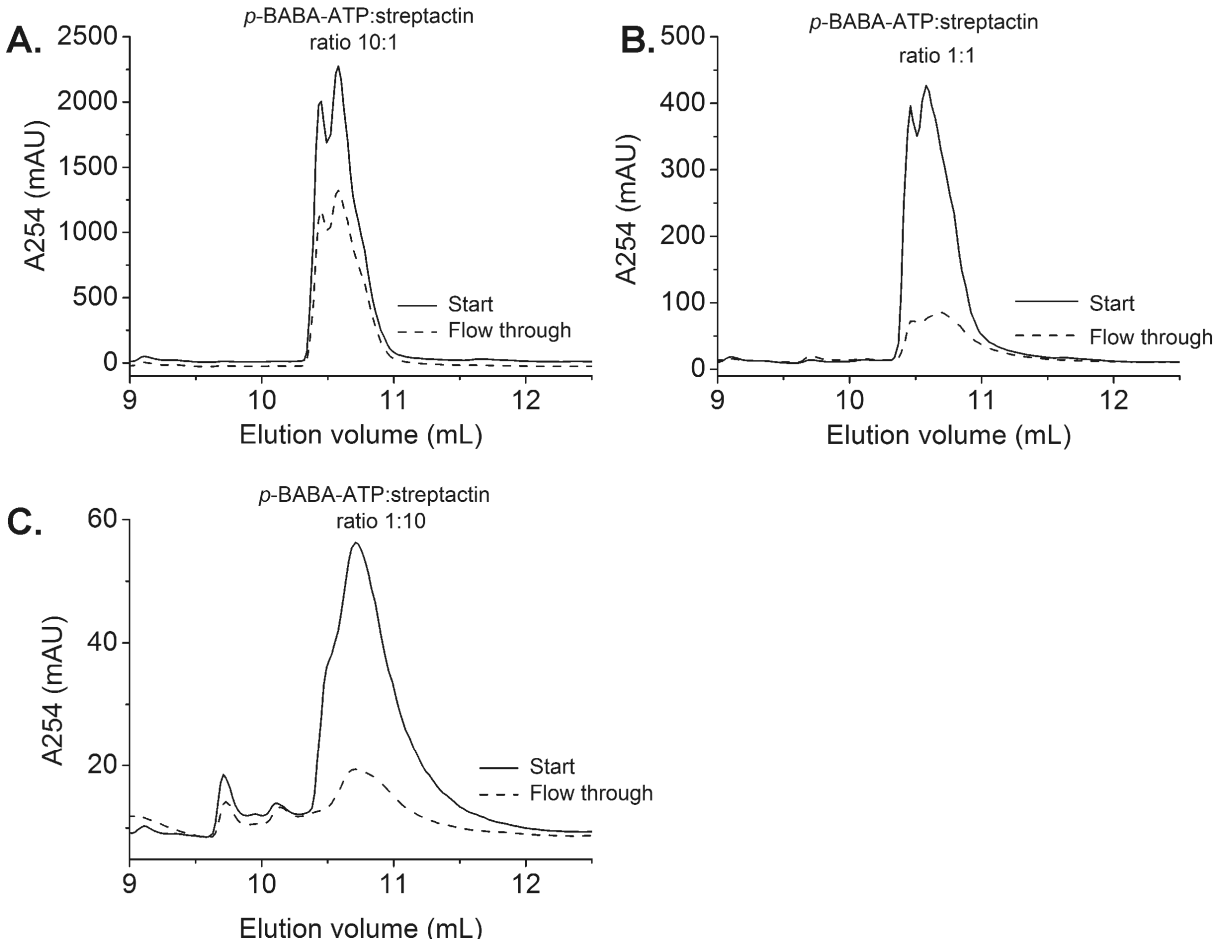


Figure 3. Analytical RP-HPLC analysis of *p*-BABA-ATP binding to the streptactin column material. The column material was incubated with different molar concentrations of *p*-BABA-ATP in the ratio's 1:10, 1:1 and 10:1, by keeping the concentration of binding sites on the column material constant at 7.5 nmol. The peak areas of the ATP analog in the RP-HPLC analysis of the starting material and of the flow through were determined.

Binding of the high-affinity nucleotide-binding protein GlcV

GlcV bound efficiently to streptactin column material in the presence of *p*-BABA-ATP as can be seen in Figure 4B (Table S1.2). Spiking of GlcV to a crude cell free extract of *L.lactis* at a protein concentration of 3 µM still yielded a highly significant binding in the presence of *p*-BABA-ATP (Figure 4C, Table S1.11). Quantification of the signals on the SDS-PAGE gel showed that about 41 % of the GlcV was bound to the column in the presence of *p*-BABA-ATP, which compares to 6 % (unspecific binding) in the absence of ATP analog. The effects of different buffer components (KCl, glycerol and DDM) in combination with KP_i showed that high KCl concentrations had a negative effect, DDM had no effect and glycerol had a positive effect. The use of Na-MES buffer instead of KP_i had a positive effect as well (Table S1.3). Next, we tested whether the negative effect of KP_i was due to the phosphate, the potassium or both. The screening clearly indicated that both the potassium (both in KP_i and KCl) and the phosphate (KP_i and NaP_i) had a negative effect on binding of GlcV to the streptactin column. Sodium ions (in NaCl) did

not affect binding and the MES-buffer showed an increased binding compared to the phosphate buffers (Table S1.4).

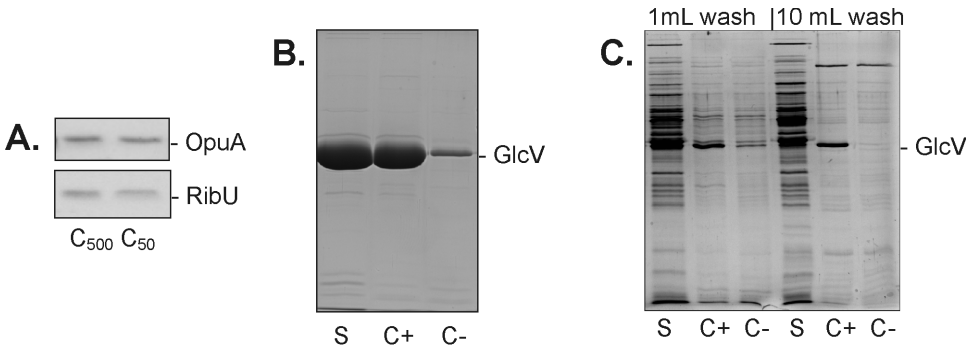


Figure 4. SDS-PAA gel and immunoblots of protein binding to streptactin columns in the presence or absence of the ATP analog *p*-BABA-ATP. The following ABC transporter proteins were used: OpuA and the soluble nucleotide-binding protein GlcV of a glucose ABC transporter from *S.solfataricus* and the non-ABC transporter protein RibU was used as control. Annotations are: start material (S) and fractions eluted from the column (C); (+) means that proteins were incubated with the column material in the presence of *p*-BABA-ATP and (–) in the absence of an ATP analog. Equal volumes of each fraction were loaded on the gel. Panel A: immunoblot of membrane vesicles containing overproduced OpuA or RibU and bound in the presence of 500 or 50 nmol *p*-BABA-ATP (Table S1.1). Panel B and C: SDS-PAA gels showing the binding of GlcV in the presence or absence of *p*-BABA-ATP as a purified fraction at a concentration of 17 µM (panel B) or 3 µM purified GlcV mixed in a *L.lactis* crude cell extract (panel C), using conditions described in Table S1.2 and S1.11.

Binding of the low-affinity ABC transporter OpuA

Binding of OpuA was tested at different stages of the purification, namely the membrane vesicles (MV), the solubilized MV (solMV) and purified protein, but no large differences were observed (Table S1.5). The binding of purified OpuA was not improved in the presence *ortho*-vanadate (together with magnesium), which is a phosphate analogue that traps ATP in the binding pocket of nucleotide-binding proteins via replacement of the released phosphate upon hydrolysis of MgATP to MgADP [54].

The binding of OpuA to the streptactin column with *p*-BABA-ATP was further analyzed using membrane vesicles with various overproduced proteins or mixtures of purified proteins (Figure 5-6, Table S1.6-S1.8). Membrane vesicles with OpuAΔSBD (here the substrate-binding domain is removed) or RibU (riboflavin transporter, as negative control) were incubated with the ATP analog and the column material (Figure 5, Table S1.6). On the SDS-PAGE gel, background binding was clearly detected, but on the basis of the immunoblot, OpuAΔSBD was bound better than RibU (18% of OpuAΔSBD versus 3% of RibU).

Another test with a mixture of membrane vesicles, containing the following his-tagged ABC transporters (OpuA, GlnPQ) and his-tagged non-ABC transporters (LacY, RibU) showed the following (Figure 5C, Table S1.7): OpuA was bound to the streptactin column but GlnPQ did not (see discussion below). The two non-ABC transporters did not bind to the streptactin column as expected.

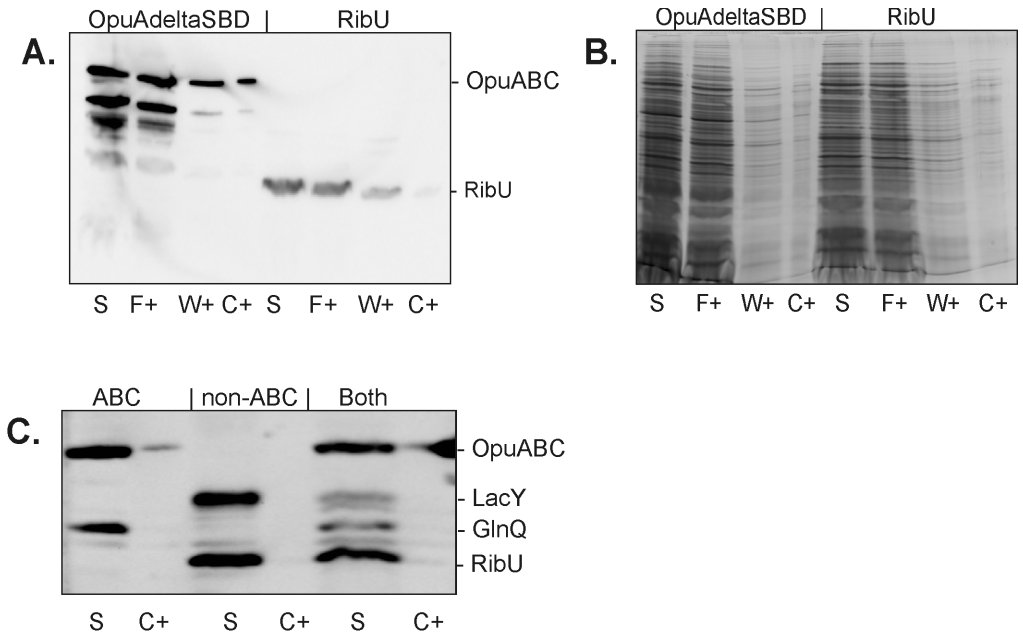


Figure 5. Immunoblots (panel A & C) and SDS-PAA gel (panel B) analysis of protein binding to streptactin columns in the presence or absence of the ATP analog *p*-BABA-ATP with membrane vesicles (1 mg/mL) overproducing OpuA△SBD or RibU (panel A & B, Table S1.6) or with a mixture of membrane vesicles (~10 mg/mL) containing various overproduced membrane proteins (panel C, Table S1.7). The following ABC transporter proteins were used: OpuA(ΔSBD) and GlnPQ and the following non-ABC transporter proteins were used as controls: RibU and LacY. Annotations are: start material (S), flow trough (F), wash fraction (W) and fractions eluted from the column (C); (+) means that proteins were incubated with the column material in the presence of *p*-BABA-ATP and (–) in the absence of an ATP analog. Equal volumes of each fraction were loaded on the gel.

Next, we compared mixtures of three ABC transporters (OpuA, GlnPQ and Sav1866) and three non-ABC transporters (LacS, RibU and IIC) (Figure 6, Table S1.8). Sav1866 and OpuA showed clear binding. The fact that OpuAA was detected by immunoblotting indicated that the whole complex of OpuA was bound since the His-tag is on OpuABC subunit and *p*-BABA-ATP is binding to the OpuAA subunit. In case of GlnPQ, the His-tag is on the ATPase subunit (GlnQ), but no binding of GlnQ was detected. However, GlnP (the transmembrane domains-substrate binding protein) was detected on SDS-PAGE gels, most likely reflecting aspecific binding. LacS and IIC did not significantly bind to the streptactin column.

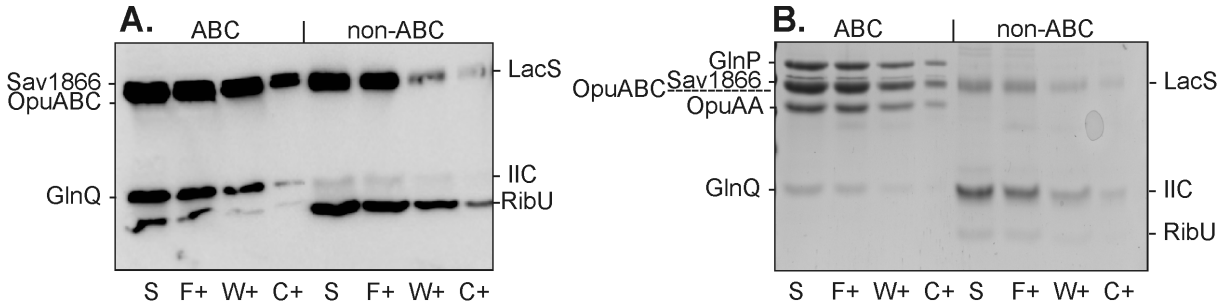


Figure 6. Immunoblot (panel A) and SDS-PAA gel (panel B) monitoring the binding of nucleotide-binding proteins from mixtures of purified proteins (0.1-0.2 mg/mL purified protein) to streptactin columns in the presence of *p*-BABA-ATP (Table S1.8). The following ABC transporter proteins were used: OpuA, GlnPQ and Sav1866. The following non-ABC transporter proteins were used as controls: RibU, LacS, and the IIC domain of the mannitol-PTS. Annotations are: start material (S), flow trough (F), wash fraction (W) and fractions eluted from the column (C); (+) means that proteins were incubated with the column material in the presence of *p*-BABA-ATP and (–) in the absence of an ATP analog. Equal volumes of each fraction were loaded on the gel.

Analysis of the binding of complex samples

The experiments with known proteins showed that *p*-BABA-ATP is able to specifically bind and enrich subsets of ABC proteins. Next, experiments were performed with complex mixtures of proteins, either soluble (cell free extract) or membrane fractions (crude membrane vesicles) of *L.lactis* and incubated with *p*-BABA-ATP. Firstly, washing of proteins bound to the streptactin column was tested with salt concentrations ranging from 50 to 500 mM NaCl (Table S1.9). Washing in terms of retaining specific binding seemed optimal between 150 and 250 mM NaCl. Secondly, Tween 20 was tested but the improvement in terms of binding was marginal (Table S1.10). Thirdly, the effect of extensive washing with MES-buffer was tested for cell free extracts spiked with GlcV (Table S1.11). Washing with 1 mL or 10 mL (10 up to 100 column volumes (CV)) did remove the unspecifically bound GlcV (in the absence of ATP analog), but also removed specific proteins from the column in the presence of ATP analog (Figure 4C). An example of the binding of proteins present in the cell free extract is shown in Figure 7 (Table S1.12). The differences between the three elution fractions were minor, indicating that most of the proteins did not bind the column via one of the two ATP analogues. Closer inspection of the SDS-PAGE gel showed small differences (Figure 7, marked with an asterisk). The marked band was sliced from the SDS-PAGE gel and identified with MALDI-MS/MSMS as a pyruvate carboxylase (pycA), an ATP-dependent carbamoyl phosphate synthase.

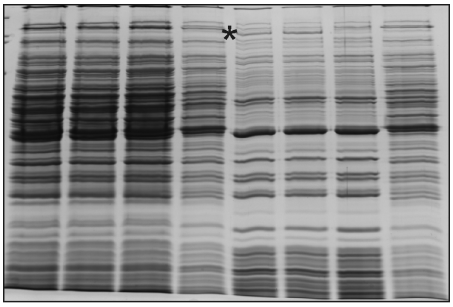


Figure 7. SDS-PAA gels showing the binding of nucleotide-binding proteins from *L.lactis* soluble cell free extract (~1 mg/mL) in the presence or absence of the ATP analog *p*-BABA-ATP or *p*-BAMBA-ATP (Table S1.12). The band marked with an asterisk was identified by MALDI-MS/MSMS. Annotations are: start material (S), flow trough (F), wash fraction (W) and fractions eluted from the column (C); (+) means that proteins were incubated with the column material in the presence of *p*-BABA-ATP; (++) in the presence of *p*-BAMBA-ATP and (–) in the absence of an ATP analog. Equal volumes of each fraction were loaded on the gel.

A gel-free proteomics approach was used for the in depth analysis of protein enrichment via the ATP analogues. Proteins bound to streptactin in the presence (representing specific and aspecific binding) or absence (respresenting aspecific binding) of *p*-BABA-ATP, were identified using mass spectrometry (after digestion of the proteins with trypsin and separation of the peptides over a RP-nLC column). Quantification was done by the iTRAQ differential labeling technique. Since, the elutions of the proteins bound to the Streptactin column with SDS loading buffer is incompatible (possibly due to the presence of staining solution and high concentrations of SDS) with the trypsin digestion and iTRAQ labeling other elution conditions were tested. Elution with MES–buffer plus 4% SDS was tested first (Table S1.13). Since the SDS concentration was well above the levels tolerated for trypsin digestion, acetone precipitation was employed to remove the SDS prior to trypsin digestion (Table S1.14). Tests with column elution fractions showed that about 75% of the sample was restored after acetone precipitation. Since a loss of 25% is substantial, other conditions were tested in an experiment to enrich for adenine nucleotide-binding proteins from DDM-solubilized membrane vesicles (with overexpressed levels of OpuA): (i) elution with acid (10% Acetic acid) (ii) with base (100 mM NaOH); and (iii) trypsin digestion directly on the column (Table S1.15). All elution conditions methods showed the presence of proteins, either on gel and/or via the UV signal in the nLC spectra (Figure S1A), but only a small number of proteins was identified after MALDI-MS/MSMS analysis, see Table S2). The largest number of proteins (13) was identified in the sample after basic elution, but more than half of the identified proteins belong to the highly abundant class of ribosomal proteins. Both subunits of the ABC transporter OpuA were also identified. Another adenine nucleotide-binding protein was the elongation factor Tu (TufA), whereas llmg_2426 is a hypothetical protein that belongs to the haloacid dehalogenase-like hydratases. The on-column digestion did not retrieve any proteins other than keratins. The various ribosomal proteins detected were more likely unspecifically bound, since they are highly abundant in the cell.

Overall, elution with base seemed most promising for detection of proteins by tryptic digestion and mass spectrometry.

Because all the peptides eluted at high acetonitril concentrations from the reverse phase liquid chromatography column, several test experiments were performed to shift the elution patterns to lower acetonitril concentrations. Clean up of the samples either with a C18 micropipet tip or with a strong cation exchange cartridge not improve the eluction for the reverse-phase column (Table S1.16). Elution of the C18 micropipet tip with methanol or acetonitril (Table S1.17) did seem to make a difference in the number of identified proteins, where elution with methanol decreased the number of proteins identified with MALDI-MS/MSMS analysis. Since cleaning of the samples did not affect the elution pattern of the peptides from the chromatography column, alterations were made in the elution gradient and fraction collection from column to optimize the separation and collection of peptides eluting at the high concentrations of acetonitril (the difference in separation using different gradient can be seen in Figure S1).

Identification and quantification of complex samples to streptactin in the presence of p-BABA-ATP

Two sets of complex samples were tested: the solubilized membrane fraction of *L.lactis* (with overexpressed OpuA) and the soluble *L.lactis* cell free extract in the presence or absence of *p*-BABA-ATP (both done in duplicate). The proteins bound on the column were eluted with base, digested with trypsin and labeled with iTRAQ label (4-plex) and combined as described in the methods section. Reverse phase liquid chromatography separation gave a clear UV signal for both fractions (UV spectra see Figure S1B)) and the collected fractions were analyzed by MALDI-MS/MSMS. Subsequent identification and quantification with MS/MSMS did not yield many identified proteins (Table 1). Only proteins that were identified with at least two peptides are listed. In the analyzed soluble fractions, OpuAA was detected together with elongation factor (tufA), the phosphopyruvate hydratase (eno) and the pyruvate carboxylase (pycA). The pyruvate carboxylase and TufA both interact with nucleotides.

Table 1. Proteins identified in the cell free extract and solubilized membrane fractions from *L.lactis*

Start material	**Accession nr**	**Protein name**	**peptide count**	**ratio 1 (114:115)**	**p-value 1 (114:115)**	**Ratio 2 (116:117)**	**p-value 2 (116:117)**
Cell free extract	llmg_0617	eno phosphopyruvate hydratase	4	2.031	0.134	1.296	0.411
	llmg_0634	pycA pyruvate carboxylase	2	1.460	0.352	1.435	0.209
	llmg_2050	tufA elongation factor Tu	3	2.344	0.081	1.240	0.251
Solubilized membrane fraction	-	OpuABC-His	2	1.256	0.530	1.130	0.403

These results suggest that adenine-nucleotide binding proteins can be enriched with the aid of the ATP analogues, but the method needs further optimization to increase the detection levels. For future experiment efforts might be focused on other methods to identify and quantify the proteins bound to the streptavidin column material. Crosslinking of the ATP analogues, using azido-groups on the adenine ring might help to reduce the background binding. To then overcome the problem of eluting the proteins, one could perform on-column trypsin digestion or introduce a cleavable tag, like a tobacco etch virus protease tag.

Material and methods

Protein constructs

The following ABC transporters were used: OpuA [55], OpuAΔSBD [56], GlnPQ [57], Sav1866 [58] and the soluble nucleotide-binding domain (GlcV) of the ABC transporter GlcSTUV [59]. The following membrane proteins were used as controls (non-ABC transporters): LacS (C320A/A635C) [60], LacY [61], IIC [62] and RibU [63]. The Sav1866 construct was cloned into the N-Lic vector pNZnLIC as described by Geertsema et al. [64]. GlcV and LacY were expressed in *E.coli,* and all other proteins in *L.lactis* as described in the listed references. GlcV was obtained from M. Pretz and C. van der Does (University Groningen, The Netherlands).

Profiling with ATP analogues

Isolation of membrane vesicles and purification of the proteins was done as described in [65]. Membrane vesicles were prepare in 50 mM KPi pH 7 at a concentration of ca 20 mg/mL, unless indicated otherwise; for protein purifications, the membrane vesicles were diluted to 5 mg/mL in 50 mM KPi pH 7, 200 mM KCl plus 20% glycerol and, subsequently, solubilized with 0.5% DDM; purified protein was obtained in 50 mM KPi pH 7, 200 mM KCl, 20% glycerol, 0.05% DDM plus 200 mM imidazole. GlcV was purified in 20 mM Na-MES pH 6.5 plus 100 mM NaCl as described by Pretz and collegues [66] and stored in aliquots at -20°C.

For profiling with the ATP analogues, protein mixtures were incubated with streptactin column material (Streptactin Sepharose from IBA, binding capacity 300 nmol/mL) in the presence of the ATP analog, using Biospin Poly-Prep columns (volume 1.2 mL, Biorad). The mixture was incubated for at least 1h at 4°C (under continuous mixing), before washing and elution of the proteins bound to the column material. The details of the buffer compositions for incubation, washing and elution are described in the figure legends. Elution of the bound proteins was done mostly with 2x SDS loading buffer, which is composed of 100 mM Tris pH 6 plus 20% glycerol, 0.002% bromophenol blue, 4 % SDS plus 1 % β-mercaptoethanol. Analysis of the protein samples was done on 12.5% SDS-PAGE gels (stained with Coomassie brilliant blue) or subsequent immunoblotting on polyvinylidene difluoride (PVDF) membrane, using the semidry electroblotting technique (blotting for 35 min at 80 mA per blot). Immunodetection was done with an antibody

(Amersham Pharmacia Biotech) against the His-tag of the various proteins. The signals were imaged with a CCD camera (Fujifilm, Las-3000 imaging system), using the Western light kit (Tropix Inc.) and quantification of the signals from the SDS-PAGE gels and immunoblots was done with the AIDA image analyzer software (version 4.15).

Trypsin digestion and iTRAQ labeling

For the iTRAQ labeling, 10 or 20 µg of total protein, originating from the fraction of the Streptactin column, were freeze-dried and resuspended in 20 µl dissolution buffer (500 mM triethyl ammonium-bicarbonate, pH 8.5) plus 0.1 % SDS. Reduction of disulfide bonds with tris-(2-carboxyethyl)phosphine (TCEP), cysteine blocking with methyl methanethiosulfonate (MMTS dissolved in isopropanol), digestion with trypsin (catalog number V5111, Promega) and iTRAQ labeling were done as described by the manufacturers protocol (Applied Biosystems) with a minor change: The iTRAQ labels were resuspended in 200 µL ethanol instead of the 70 µl ethanol described in the labeling protocol, to maintain the high percentage of organic solvent during labeling. This modification was made because the iTRAQ labels were divided over two different samples (thus, 100 µL was used per reaction). The 4-plex labels were used to label the duplicate samples of protein elution fractions from streptactin columns with (114 and 116) and without ATP analog (115 and 117) to determine their relative enrichment.

Reverse phase nano-liquid chromatography (RP-nLC) and MALDi-MS/MSMS analysis

For the separation on RP-nLC, peptides were trapped on a pre-column (300 µm x 5 mm, 5 µm particles PepMap C18, p/n 160454) and subsequently separated over an analytical C18 column (75 µm x 150 mm, 3 µm particles 100 Å pores, Acclaim PepMap100 C18, p/n 160321), using the Ultimate 3000 nano LC system (Dionex). The gradient solutions contained 0.05% TFA (solution A) plus 80% ACN plus 0.05% TFA (solution B), and the gradient was run at a flow rate of 300 nl/min. Equilibration of the column, binding and washing of the peptides was performed with 4% solution B. Two different gradients were used for the elution of the peptides: the 'normal' gradient was ran with 4-40% solution B in 50 min, 40-60% solution B in 10 minutes and 60-100% solution B in 5 min. Collection of the eluted peptides was started after 15 minutes, and the fractions were mixed 1:4 with 2.4 mg/ml α-cyano-4-hydroxy-cinnamic acid matrix (LaserBio Labs) plus an internal calibration mixture (3 fmol/µL angiotensin II (Sigma-Aldrich) and 6 fmol/µL adrenocorticotropic hormone (ACTH) fragment 18-39 (Sigma-Aldrich)) and spotted directly onto a MALDI target. Fractions of 12 seconds were spotted (260 or 520 spots per nLC run), using the Probot system (LC Packings, Amsterdam, The Netherlands). The second gradient used on the RP-nLC column ('linear' gradient) was setup because peptides only eluted at high ACN concentrations. The elution was run from 4-100% solution B in 95 min and, subsequently, 100% solution B in 20 min. For these experiments, the Probot was started after 5 min, and 520 fractions of 12 seconds each were spotted.

Peptides were analyzed with a 4700 or 4800 MALDI-TOF/TOF Analyzer (Applied Biosystems, Foster City, CA – USA). The MALDI-TOF/TOF was operated in reflectron positive ionization mode in the m/z range 700-4000. MS/MS precursors were selected from the top 30 peaks above the signal-to-noise (S/N) threshold of 50 per MS spectra. The MS/MS spectra were acquired using a 2kV acceleration voltage and air as collision gas at $1x10^{-6}$ torr. A peak list of the acquired MS/MS spectra was generated using default settings. The MS spectra were calibrated using internal calibration of angiotensin II (m/z = 1046.5420) and ACTH 18-39 (m/z = 2465.1990).

Cleaning of the samples

Strong cation exchange (SCX) cartridges: Part of the trypsin digested sample (after digestion the samples were freeze-dried and stored as powder at -20°C) was resuspended in 10 µL of 10 mM triethyl ammonium phosphate pH 2.7 plus 25% ACN (buffer S1) and the pH was set to 2.5 by additions from a stock of 1M H_2PO_4. The SCX cartridge (TopTip PolySulfoethyl A, Glygen corp.) was pre-equilibrated in buffer S1; the sample was pipetted two times over the tip, which was then washed 3 times with buffer S1 and eluted with 10 µL buffer S1 plus 500 mM KCl by pipetting the solution multiple times over the tip. All fractions were collected for MS analysis or further cleaning using a C18 micropipet tip.

C18 micropipet tip: Samples were diluted in 20 µL 0.1% TFA (samples from the SCX cartridges were diluted at least 10-fold to reduce the ACN concentration). Samples were pipetted up and down over the pre-equilibrated micropipet tip to bind the samples to the C18 column material; the micropipette tip was washed with 0.1% TFA and eluted in three steps of 5 µL, using increasing ACN concentrations of 35%, 50% and 70%. These fractions could be analyzed separately on MS/MSMS or could be combined for further separation on RP-nLC.

Analytical RP-HPLC

Analytical RP-HPLC was performed using a polystyrene/divenyl benzene column (Source^{TM} 5RPC ST 4.6/150) connected to an Äkta-purifier system. Two buffers were used for the separation: Buffer A2: 100 mM triethyl ammonium acetate (TEAA) pH 7.0 and Buffer B2: acetonitrile plus 100 mM TEAA pH 7.0. The settings for the separations were: 200 µL loop, flow rate 1 mL/min, gradient: 5 min 17% buffer B2, 10 min gradient 17-20% buffer B2, 5 min 90% buffer B2, 5 min 0%B.

Conclusion

In Chapter 2, the synthesis of two ATP analogues (*p*-BABA-ATP and *p*-BAMBA-ATP) was described, for ABC-type nucleotide-binding proteins had an affinity at least one order of magnitude higher than for ATP. The presence of the biotin moiety on the analogues was exploited for the enrichment and profiling of nucleotide-binding proteins. Relatively specific binding was demonstrated for several ABC transporters, but in most cases the

affinity of the analogues was still too low to discriminate effectively between nucleotide-binding and non-nucleotide binding proteins.

Acknowledgements

We would like to thank the following people for providing purified proteins or membrane vesicles: Monika Pretz and Chris van der Does (purified GlcV), Gea Schuurman-Wolters (GlnPQ, IIC, Sav1866, LacY), Siva Ramadurai (LacS), Josy ter Beek (RibU), Guus Erkens (soluble cell free extracts) and Faizah Fulyani (soluble cell free extracts). We would like to thank Wim Huibers and Fabrizia Fusetti for their help with the LC-MS measurements. This work was supported by grants from The Netherlands Organisation for Scientific research (NWO, Chemical Sciences Top Subsidy to BP; grant number 700-56-302 and the Netherlands Proteomics Centre (NPC).

References

[1] Heal,W.P., Wickramasinghe,S.R., & Tate,E.W. (2008) Activity based chemical proteomics: profiling proteases as drug targets. *Curr. Drug Discov. Technol.*, **5,** 200-212.

[2] Barglow,K.T. & Cravatt,B.F. (2007) Activity-based protein profiling for the functional annotation of enzymes. *Nat. Methods*, **4,** 822-827.

[3] Cravatt,B.F., Wright,A.T., & Kozarich,J.W. (2008) Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev Biochem*, **77,** 383-414.

[4] Evans,M.J. & Cravatt,B.F. (2006) Mechanism-based profiling of enzyme families. *Chem Rev*, **106,** 3279-3301.

[5] Hagenstein,M.C. & Sewald,N. (2006) Chemical tools for activity-based proteomics. *J Biotechnol*, **124,** 56-73.

[6] Schmidinger,H., Hermetter,A., & Birner-Gruenberger,R. (2006) Activity-based proteomics: enzymatic activity profiling in complex proteomes. *Amino Acids*, **30,** 333-350.

[7] Sieber,S.A. & Cravatt,B.F. (2006) Analytical platforms for activity-based protein profiling--exploiting the versatility of chemistry for functional proteomics. *Chem Commun (Camb.)*, 2311-2319.

[8] Uttamchandani,M., Li,J., Sun,H., & Yao,S.Q. (2008) Activity-based protein profiling: new developments and directions in functional proteomics. *Chembiochem.*, **9,** 667-675.

[9] Jessani,N., Niessen,S., Wei,B.Q., Nicolau,M., Humphrey,M., Ji,Y., Han,W., Noh,D.Y., Yates,J.R., III, Jeffrey,S.S., & Cravatt,B.F. (2005) A streamlined platform for high-content functional proteomics of primary human specimens. *Nat. Methods*, **2,** 691-697.

[10] Speers,A.E. & Cravatt,B.F. (2005) A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *J Am. Chem Soc*, **127,** 10018-10019.

[11] Kolb,H.C., Finn,M.G., & Sharpless,K.B. (2001) Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem Int Ed Engl.*, **40,** 2004-2021.

[12] Speers,A.E., Adam,G.C., & Cravatt,B.F. (2003) Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J Am. Chem Soc*, **125,** 4686-4687.

[13] Speers,A.E. & Cravatt,B.F. (2004) Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol*, **11,** 535-546.

[14] Lo,L.C., Pang,T.L., Kuo,C.H., Chiang,Y.L., Wang,H.Y., & Lin,J.J. (2002) Design and synthesis of class-selective activity probes for protein tyrosine phosphatases. *J Proteome. Res*, **1,** 35-40.

[15] Zhu,Q., Huang,X., Chen,G.Y.J., & Yao,S.Q. (2003) Activity-based fluorescent probes that target phosphatases. *Tetrahedron Letters*, **44,** 2669-2672.

[16] Tsai,C.S., Li,Y.K., & Lo,L.C. (2002) Design and synthesis of activity probes for glycosidases. *Org. Lett*, **4,** 3607-3610.

[17] Zhu,Q., Girish,A., Chattopadhaya,S., & Yao,S.Q. (2004) Developing novel activity-based fluorescent probes that target different classes of proteases. *Chem Commun (Camb.)*, 1512-1513.

[18] Williams,S.J., Hekmat,O., & Withers,S.G. (2006) Synthesis and testing of mechanism-based protein-profiling probes for retaining endo-glycosidases. *Chembiochem.*, **7,** 116-124.

[19] Kurogochi,M., Nishimura,S., & Lee,Y.C. (2004) Mechanism-based fluorescent labeling of beta-galactosidases. An efficient method in proteomics for glycoside hydrolases. *J Biol Chem*, **279,** 44704-44712.

[20] Liu,Y., Patricelli,M.P., & Cravatt,B.F. (1999) Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci U. S. A*, **96,** 14694-14699.

[21] Kidd,D., Liu,Y., & Cravatt,B.F. (2001) Profiling serine hydrolase activities in complex proteomes. *Biochemistry*, **40,** 4005-4015.

[22] Pan,Z., Jeffery,D.A., Chehade,K., Beltman,J., Clark,J.M., Grothaus,P., Bogyo,M., & Baruch,A. (2006) Development of activity-based probes for trypsin-family serine proteases. *Bioorg. Med. Chem Lett*, **16,** 2882-2885.

[23] Gillet,L.C., Namoto,K., Ruchti,A., Hoving,S., Boesch,D., Inverardi,B., Mueller,D., Coulot,M., Schindler,P., Schweigler,P., Bernardi,A., & Gil-Parrado,S. (2008) In-cell selectivity profiling of serine protease inhibitors by activity-based proteomics. *Mol Cell Proteomics*, **7,** 1241-1253.

[24] Schmidinger,H., Birner-Gruenberger,R., Riesenhuber,G., Saf,R., Susani-Etzerodt,H., & Hermetter,A. (2005) Novel fluorescent phosphonic acid esters for discrimination of lipases and esterases. *Chembiochem.*, **6,** 1776-1781.

[25] Kumar,S., Zhou,B., Liang,F., Wang,W.Q., Huang,Z., & Zhang,Z.Y. (2004) Activity-based probes for protein tyrosine phosphatases. *Proc. Natl. Acad. Sci U. S. A*, **101,** 7943-7948.

[26] Kumar,S., Zhou,B., Liang,F., Yang,H., Wang,W.Q., & Zhang,Z.Y. (2006) Global analysis of protein tyrosine phosphatase activity with ultra-sensitive fluorescent probes. *J Proteome. Res*, **5,** 1898-1905.

[27] Liu,S., Zhou,B., Yang,H., He,Y., Jiang,Z.X., Kumar,S., Wu,L., & Zhang,Z.Y. (2008) Aryl vinyl sulfonates and sulfones as active site-directed and mechanism-based probes for protein tyrosine phosphatases. *J Am. Chem Soc*, **130,** 8251-8260.

[28] Palmer,J.T., Rasnick,D., Klaus,J.L., & Bromme,D. (1995) Vinyl sulfones as mechanism-based cysteine protease inhibitors. *J Med. Chem*, **38,** 3193-3196.

[29] Wang,G., Mahesh,U., Chen,G.Y., & Yao,S.Q. (2003) Solid-phase synthesis of peptide vinyl sulfones as potential inhibitors and activity-based probes of cysteine proteases. *Org. Lett*, **5,** 737-740.

[30] Borodovsky,A., Ovaa,H., Kolli,N., Gan-Erdene,T., Wilkinson,K.D., Ploegh,H.L., & Kessler,B.M. (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem Biol*, **9,** 1149-1159.

[31] Nazif,T. & Bogyo,M. (2001) Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc. Natl. Acad. Sci U. S. A*, **98,** 2967-2972.

[32] Greenbaum,D., Medzihradszky,K.F., Burlingame,A., & Bogyo,M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem Biol*, **7,** 569-581.

[33] van Swieten,P.F., Maehr,R., van den Nieuwendijk,A.M., Kessler,B.M., Reich,M., Wong,C.S., Kalbacher,H., Leeuwenburgh,M.A., Driessen,C., van der Marel,G.A., Ploegh,H.L., & Overkleeft,H.S. (2004) Development of an isotope-coded activity-based probe for the quantitative profiling of cysteine proteases. *Bioorg. Med. Chem Lett*, **14,** 3131-3134.

[34] Liau,M.L., Panicker,R.C., & Yao,S.Q. (2003) Design and synthesis of an affinity probe that targets caspases in proteomic experiments. *Tetrahedron Letters*, **44,** 1043-1046.

[35] Kato,D., Boatright,K.M., Berger,A.B., Nazif,T., Blum,G., Ryan,C., Chehade,K.A., Salvesen,G.S., & Bogyo,M. (2005) Activity-based probes that target diverse cysteine protease families. *Nat. Chem Biol*, **1,** 33-38.

[36] Thornberry,N.A., Peterson,E.P., Zhao,J.J., Howard,A.D., Griffin,P.R., & Chapman,K.T. (1994) Inactivation of interleukin-1 beta converting enzyme by peptide (acyloxy)methyl ketones. *Biochemistry*, **33,** 3934-3940.

[37] Bottcher,T. & Sieber,S.A. (2008) Beta-lactones as privileged structures for the active-site labeling of versatile bacterial enzyme classes. *Angew. Chem Int Ed Engl.*, **47,** 4600-4603.

[38] Staub,I. & Sieber,S.A. (2008) Beta-lactams as selective chemical probes for the in vivo labeling of bacterial enzymes involved in cell wall biosynthesis, antibiotic resistance, and virulence. *J Am. Chem Soc*, **130,** 13400-13409.

[39] Wright,A.T. & Cravatt,B.F. (2007) Chemical proteomic probes for profiling cytochrome p450 activities and drug interactions in vivo. *Chem Biol*, **14,** 1043-1051.

[40] Khandekar,S.S., Feng,B., Yi,T., Chen,S., Laping,N., & Bramson,N. (2005) A liquid chromatography/mass spectrometry-based method for the selection of ATP competitive kinase inhibitors. *J Biomol. Screen.*, **10,** 447-455.

[41] Li,Y.M., Xu,M., Lai,M.T., Huang,Q., Castro,J.L., DiMuzio-Mower,J., Harrison,T., Lellis,C., Nadin,A., Neduvelil,J.G., Register,R.B., Sardana,M.K., Shearman,M.S., Smith,A.L., Shi,X.P., Yin,K.C., Shafer,J.A., & Gardell,S.J. (2000) Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature*, **405,** 689-694.

[42] Salisbury,C.M. & Cravatt,B.F. (2007) Activity-based probes for proteomic profiling of histone deacetylase complexes. *Proc. Natl. Acad. Sci. USA*, **104,** 1171-1176.

[43] Salisbury,C.M. & Cravatt,B.F. (2008) Optimization of activity-based probes for proteomic profiling of histone deacetylase complexes. *J. Am. Chem. Soc*, **130,** 2184-2194.

[44] Saghatelian,A., Jessani,N., Joseph,A., Humphrey,M., & Cravatt,B.F. (2004) Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. USA*, **101,** 10000-10005.

[45] Patricelli,M.P., Szardenings,A.K., Liyanage,M., Nomanbhoy,T.K., Wu,M., Weissig,H., Aban,A., Chun,D., Tanner,S., & Kozarich,J.W. (2007) Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry*, **46,** 350-358.

[46] Zheng,J., Knighton,D.R., Ten Eyck,L.F., Karlsson,R., Xuong,N., Taylor,S.S., & Sowadski,J.M. (1993) Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry*, **32,** 2154-2161.

[47] Hanks,S.K. & Hunter,T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J*, **9,** 576-596.

[48] Adam,G.C., Cravatt,B.F., & Sorensen,E.J. (2001) Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem. Biol*, **8,** 81-95.

[49] Adam,G.C., Sorensen,E.J., & Cravatt,B.F. (2002) Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes. *Mol Cell Proteomics*, **1,** 828-835.

[50] Adam,G.C., Sorensen,E.J., & Cravatt,B.F. (2002) Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol*, **20,** 805-809.

[51] Barglow,K.T. & Cravatt,B.F. (2004) Discovering disease-associated enzymes by proteome reactivity profiling. *Chem Biol*, **11,** 1523-1531.

[52] Davidson,A.L., Dassa,E., Orelle,C., & Chen,J. (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev*, **72,** 317-64, table.

[53] Dean,M. (2005) The genetics of ATP-binding cassette transporters. *Methods Enzymol*, **400,** 409-429.

[54] Goodno,C.C. (1982) Myosin active-site trapping with vanadate ion. *Methods Enzymol*, **85 Pt B,** 116-123.

[55] van der,H.T. & Poolman,B. (2000) Osmoregulated ABC-transport system of Lactococcus lactis senses water stress via changes in the physical state of the membrane. *Proc. Natl. Acad. Sci U. S. A*, **97,** 7102-7106.

[56] Biemans-Oldehinkel,E. & Poolman,B. (2003) On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. *EMBO J*, **22,** 5983-5993.

[57] Schuurman-Wolters,G.K. & Poolman,B. (2005) Substrate specificity and ionic regulation of GlnPQ from Lactococcus lactis. An ATP-binding cassette transporter with four extracytoplasmic substrate-binding domains. *J Biol Chem*, **280,** 23785-23790.

[58] Dawson,R.J. & Locher,K.P. (2006) Structure of a bacterial multidrug ABC transporter. *Nature*, **443,** 180-185.

[59] Verdon,G., Albers,S.V., Dijkstra,B.W., Driessen,A.J., & Thunnissen,A.M. (2002) Purification, crystallization and preliminary X-ray diffraction analysis of an archaeal ABC-ATPase. *Acta Crystallogr D Biol Crystallogr*, **58,** 362-365.

[60] Doeven,M.K., Folgering,J.H., Krasnikov,V., Geertsma,E.R., van den,B.G., & Poolman,B. (2005) Distribution, lateral mobility and function of membrane proteins incorporated into giant unilamellar vesicles. *Biophys J*, **88,** 1134-1142.

[61] Teather,R.M., Bramhall,J., Riede,I., Wright,J.K., Furst,M., Aichele,G., Wilhelm,U., & Overath,P. (1980) Lactose carrier protein of Escherichia coli. Structure and expression of plasmids carrying the Y gene of the lac operon. *Eur. J Biochem*, **108,** 223-231.

[62] Boer,H., Hoeve-Duurkens,R.H., Schuurman-Wolters,G.K., Dijkstra,A., & Robillard,G.T. (1994) Expression, purification, and kinetic characterization of the mannitol transport domain of the phosphoenolpyruvate-dependent mannitol phosphotransferase system of Escherichia coli. Kinetic evidence that the E. coli mannitol transport protein is a functional dimer. *J Biol Chem*, **269,** 17863-17871.

[63] Burgess,C.M., Slotboom,D.J., Geertsma,E.R., Duurkens,R.H., Poolman,B., & van Sinderen,D. (2006) The riboflavin transporter RibU in Lactococcus lactis: molecular characterization of gene expression and the transport mechanism. *J Bacteriol*, **188,** 2752-2760.

[64] Geertsma,E.R. & Poolman,B. (2007) High-throughput cloning and expression in recalcitrant bacteria. *Nat. Methods*, **4,** 705-707.

[65] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. USA*, **103,** 10624-10629.

[66] Pretz,M.G., Albers,S.V., Schuurman-Wolters,G., Tampe,R., Driessen,A.J.M., & van der Does,C. (2006) Thermodynamics of the ATPase cycle of GlcV, the nucleotide-binding domain of the glucose ABC transporter of Sulfolobus solfataricus. *Biochemistry*, **45,** 15056-15067.

Chapter 4 Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA

Justina C. Wolters [#], Ronnie P-A. Berntsson [#], Nadia Gul, Akira Karasawa, Andy-Mark W. H. Thunnissen, Dirk-Jan Slotboom and Bert Poolman

[#] These authors contributed equally to the work

Part of this chapter is published in PLoS One. 2010 Apr 29; 5(4):e10361

Abstract

Background

The ABC transporter OpuA from *Lactococcus lactis* transports glycine betaine upon activation by threshold values of ionic strength. In this study, the ligand binding characteristics of purified OpuA in a detergent-solubilized state and of its substrate-binding domain produced as soluble protein (OpuAC) was characterized.

Principal findings

The binding of glycine betaine to purified OpuA and OpuAC (K_D = 4-6 μM) did not show any salt dependence or cooperative effects, in contrast to the transport activity. Also nucleotide binding to purified OpuA was not affected by the ionic strength. OpuAC is highly specific for glycine betaine and the related proline betaine. Other compatible solutes like proline and carnitine bound with affinities that were 3 to 4 orders of magnitude lower. The low affinity substrates were not noticeably transported by membrane-reconstituted OpuA. OpuAC was crystallized in an open (1.9 Å) and closed-liganded (2.3 Å) conformation. The binding pocket is formed by three tryptophans (Trp-prism) coordinating the quaternary ammonium group of glycine betaine in the closed-liganded structure. Even though the binding site of OpuAC is identical to that of its *B. subtilis* homolog, the affinity for glycine betaine is 4-fold higher.

Conclusions

Ionic strength did not affect nucleotide or substrate binding to OpuA, indicating that regulation of transport is not at the level of substrate binding, but rather at the level of translocation. The overlap between the crystal structures of OpuAC from *L.lactis* and *B.*subtilis, comprising the classical Trp-prism, show that the differences observed in the binding affinities originate from outside of the ligand binding site.

Introduction

The osmoregulatory ABC transporter OpuA protects *Lactococcus lactis* against hyperosmotic stress by accumulating the compatible solute glycine betaine. It has been shown that osmotic activation of OpuA depends on three factors [1-3]: (i) the osmotic signal, associated with a change in the intracellular ionic strength; (ii) the membrane

lipid composition, i.e. osmotic regulation requires threshold levels of anionic lipids; and (iii) the presence of tandem CBS domains ('CBS module') in OpuA that acts as osmosensor. Above threshold levels of anionic lipids and below the threshold ionic strength, the transporter is locked in an 'off' state, presumably via an interaction of the CBS module with the membrane surface. When the ionic strength is increased above the threshold or, alternatively, the negative surface charge of the membrane is decreased [1,3,4], the transporter is activated ('on' state). Because ionic strength and a negative surface charge act reciprocally, that is, the higher the fraction of anionic lipids - the higher the ionic strength needed for activation, it is thought that ions screen the electrostatic interaction of the CBS module with the membrane (as depicted in Fig. 1). We have recently shown that deletion of the CBS module (OpuAΔCBS mutant; Fig. 1A) or substitution of five surface-exposed cationic residues on the CBS module to neutral amino acids (OpuAK3R2 mutant) suffices for deregulated transport [5]. The OpuAΔCBS and OpuAK3R2 mutants are no longer osmotically regulated but otherwise fully functional in transport. The importance of the CBS module in osmoregulation has also been shown for an OpuA homolog in *Pseudomonas syringae* [6]. The mechanism of osmotic activation of OpuA bears resemblance to that of ProP from *Escherichia coli* [7]and BetP from *Corynebacterium glutamicum* [8], which are secondary transporters that use electrochemical ion gradients rather than ATP to drive transport. Whereas OpuA responds to ionic strength [2], BetP is specifically activated by K^+ ions [9]. ProP also responds to ionic strength but additional factors such as hydration or macromolecular crowding and recently the membrane potential have been implicated in the osmotic regulation of transport [7,10]. Studies on the transcriptional regulator of OpuA showed that the repressor of the *opuA* operon dissociates from the DNA at high ionic strengths, which is consistent with increased expression of the transporter and thus increased accumulation of glycine betaine at high osmotic stress [11].

The ABC transporter OpuA from *Lactococcus lactis* is a dimer and each half has two subunits. One subunit is composed of the nucleotide-binding domain (NBD) fused N-terminally to a tandem pair of CBS domains. The other subunit contains the transmembrane domain (TMD) fused N-terminally to the substrate-binding domain (SBD). The NBD-CBS and SBD-TMD subunits are named OpuAA and OpuABC, respectively (Fig. 1). In case of *L. lactis* OpuA, the NBD-CBS domains are followed by a stretch of about twenty amino acids, most of which are anionic. This anionic tail varies in length among OpuA homologues. It is lacking in the OpuA orthologue from *E. coli* (ProU), but can reach lengths of more than hundred amino acids long in some orthologues present in Archaea. For OpuA from *L. lactis* it has been shown that the anionic tail tunes the ionic regulation [2].

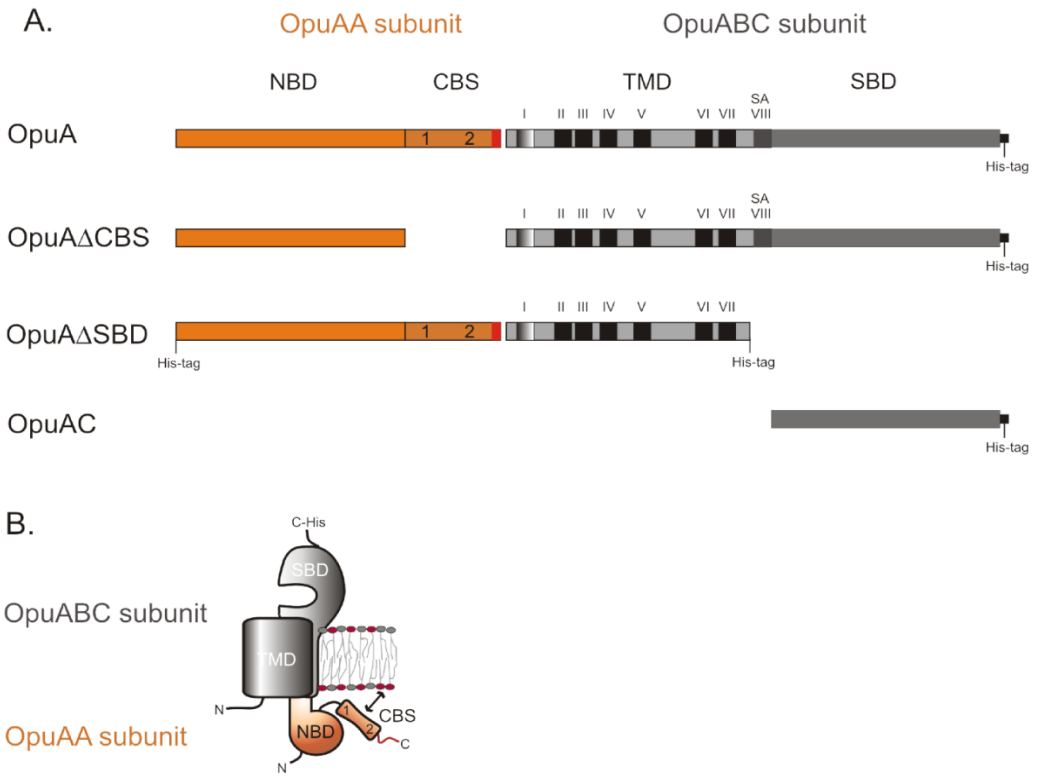


Figure 1. Schematic of OpuA from *L. lactis*. Panel A represents the different constructs used in this paper. OpuAA consists of the nucleotide-binding domain and tandem CBS domains connected to the anionic tail (approximately 20 amino acids, indicated in red). OpuABC consists of the transmembrane domain (TMD, OpuAB: I is amphipatic helix; II-VII are transmembrane segments; SA is the signal anchor sequence), connected to the substrate-binding domain (SBD, OpuAC). Panel B shows the organization of the subunits in the membrane. The functional protein is a dimer of the complex shown in panel B. The OpuABC subunits are depicted in grey, the OpuAA subunits in orange.

The SBD of OpuA, hereafter named OpuAC, belongs to a superfamily of proteins associated with ABC transporters involved in solute uptake in prokaryotes [12]. These proteins consist of two globular domains with a α/β fold that are connected by a flexible hinge. Relative movements of the domains about the hinge allow the proteins to adopt closed and open conformations. Substrates bind between the two domains and shift the equilibrium towards a closed state, a process often referred to as a Venus fly-trap mechanism [13]. The closed, ligand-bound proteins associate with the transmembrane domains and deliver the cargo for translocation. Previously, it was shown that OpuAC could complement OpuAΔSBD, albeit poorly due to the low affinity of OpuAC for the membrane-domain of OpuA [14]. OpuAC from *Bacillus subtilis* has previously been investigated in terms of ligand binding as well as having its structure determined [15,16]. This protein is membrane-tethered via a N-terminal lipid modification rather than covalently linked to the translocator domain as in OpuA from *L. lactis.* We have now characterized ligand binding to the full-length transporter OpuA and to isolated OpuAC (see Fig. 1A). Furthermore, we have determined the crystal structures of OpuAC in the absence and presence of its natural ligand, glycine betaine, at a resolution of 1.9 Å and 2.3 Å, respectively.

Results

Glycine betaine binding

Binding of glycine betaine to purified OpuA (solubilized in DDM) and OpuAC was assessed under different conditions relevant for the osmoregulatory function of the transporter. Glycine betaine binding was measured with three different methods: (i) intrinsic protein fluorescence; (ii) isothermal titration calorimetry; and (iii) a filter-based assay using radioactive ligand. Measurements were performed at different ionic strengths, using increasing concentrations of potassium phosphate (ranging from 10-250 mM KP_i, which is equivalent to an ionic strength of 0.02-0.5. The intrinsic protein fluorescence of OpuA and OpuAC decreased by the addition of glycine betaine until saturation was achieved. The maximal fluorescence change was about 9-11% for OpuA (Fig. 2A, Table 1) and 23-24% for OpuAC (Fig. 2B, Table 1). These differences reflect the larger number of tryptophan residues in OpuA (16 in total of which 10 are present in the OpuAC domain). The dissociation constants for glycine betaine binding to OpuA (K_d's around 5-6 µM, Fig. 2A, Table 1) and OpuAC (K_d's around 4-5 µM, Fig. 2B, Table 1) were similar and did not show any salt dependence. At very high protein concentration, the amount of glycine betaine bound was 0.74 mol/mol OpuAC (Fig. 2C). Although fluorescence measurements report binding constants over a wide concentration range, a K_D of ~ 5 µM is generally too high for an accurate estimate of the number of binding sites from the intercept of the slope and the plateau of the binding curve [17]. We then determined the number of sites from the binding of radiolabeled glycine betaine to purified OpuA via ammonium-sulfate precipitation [18]. Again, the dissociation constant was shown to be independent of the ionic strength (Fig. 2D). The K_d was averaged from duplicate measurements for all salt concentrations, and was 7.3 µM; the maximal number of binding sites was ~3 nmol/mg OpuA (which equals to 1.2 mol of substrate per mol OpuA or 0.6 mol per mol of substrate binding domain). To ascertain that glycine betaine binding only occurs to the OpuAC domain and not to some intracellular regulatory site, as has been observed for other ABC transporters [19,20], ligand binding to OpuAΔSBD and OpuAΔCBS was also tested. Deletion of the CBS domains did not alter the dissociation constant and OpuAΔSBD did not show any glycine betaine binding (data not shown). In principle, ligand binding could occur to the membrane domain as suggested, for instance, by the crystal structures of full length transporters [21]. However, this binding would most likely be of low affinity and not readily detected. Our studies indicate that high affinity (K_D in the low µM range) only occurs to the extracytoplasmic substrate-binding domains of OpuA.

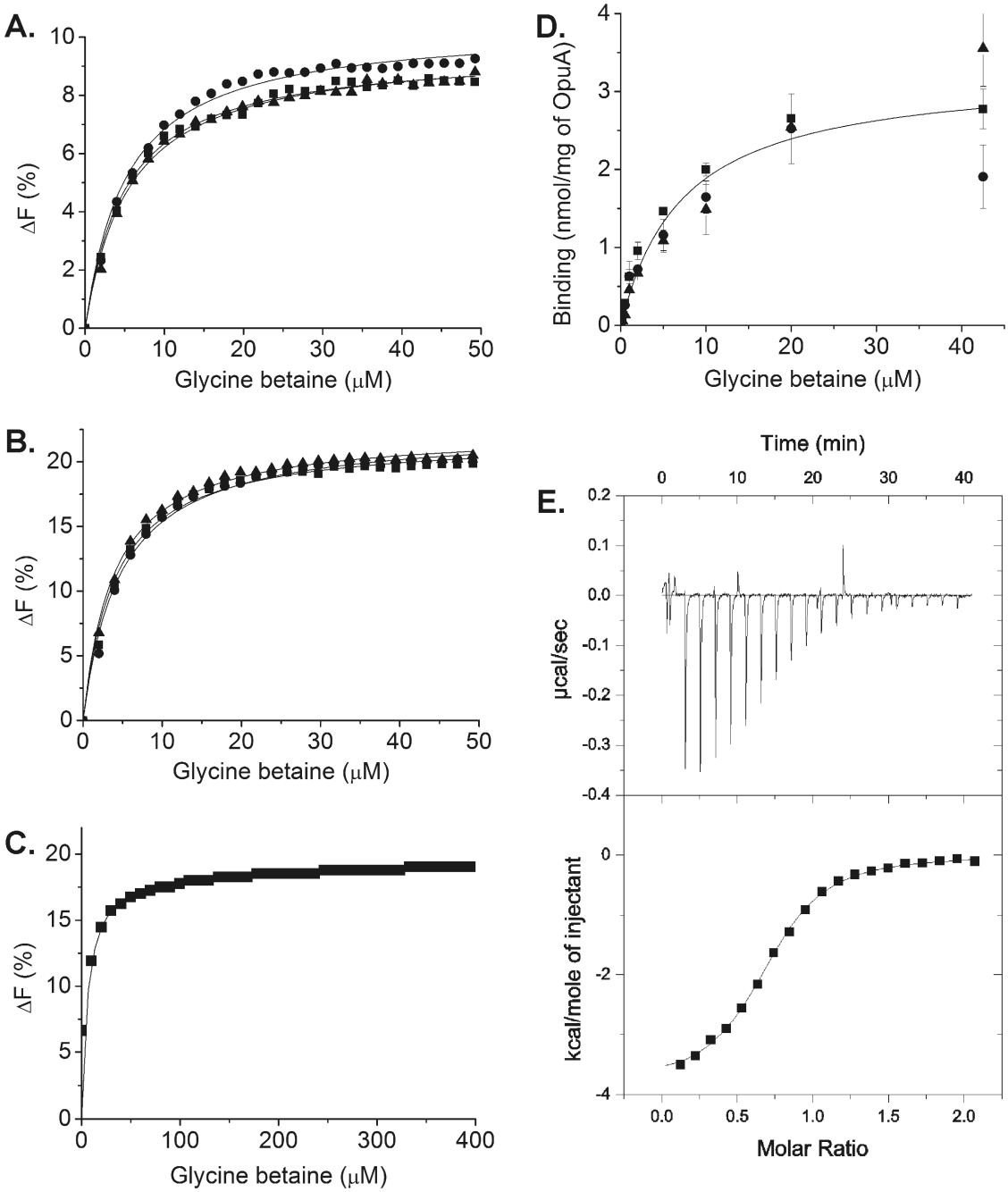


Figure 2. Glycine betaine-binding to purified OpuA (panel A & D) or OpuAC (panel B, C & E). Binding was measured using intrinsic protein fluorescence (panel A, B & C), filter-based assay using radio-labeled ligand (panel D), and isothermal titration calorimetry (panel E) at various concentrations of potassium phosphate (10-250 mM KP_i, pH 7.0). The intrinsic protein fluorescence measurements and the filter-based assay were done using a concentration of ~0.5 µM purified OpuAC (panel A, B and D) for determination of the dissociation constants and the determination of the number of binding sites (panel C) was measured at a concentration of 20 µM. OpuAC The fluorescent measurements were plotted as the absolute percentage change of the initial fluorescence signal in the absence of substrate (panel A, B & C); the filter-based measurements were corrected for background signals (panel D); and the data fits are shown as black lines. In panel A,B & D: measurements were done at 250 mM KP_i (■), 50 mM KP_i (●) and 10 mM KP_i (panel A and D) or 12.5 mM KP_i (panel B) (▲). ITC measurements (panel E) were performed at 50 µM of purified OpuAC and the integrated heat peaks were fitted to a one site-binding model (black line, lower panel).

Table 1. Glycine betaine binding to purified OpuA and OpuAC using intrinsic protein fluorescence. Dissociation constants (K_D) and maximal fluorescence change (ΔF_{max}) values of glycine betaine binding to OpuA and OpuAC at different salt concentrations. The dissociation constants and ΔF_{max} values were averaged from at least two independent measurements and the errors indicate the range.

KPi, pH7.0 (mM)	OpuA		OpuAC	
	K_D (µM)	ΔF_{max} (%)	K_D (µM)	ΔF_{max} (%)
12.5	6.0±0.5	9.9±0.3	4.5±0.7	23.6±1.1
50	6.0±0.7	10.8±0.4	4.6±0.1	23.4±0.9
250	5.2±0.2	9.3±0.2	4.1±0.0	23.1±1.2

Thermodynamic analysis of ligand binding

Isothermal titration calorimetry (ITC) was used to determine the thermodynamic parameters contributing to the ligand binding. The heat changes upon ligand binding do not only yield information on the dissociation constant (and thus ΔG) and the number of binding sites, but also on the enthalpy (ΔH) and entropy (ΔS) changes. A representative measurement, corrected for blanks, is shown in Fig. 2E, which yielded a dissociation constant of 3.8 ± 1.1 µM and 0.7 mol of glycine betaine bound per mol of OpuAC. The enthalpy change was -3.8 ± 0.1 kcal mol^{-1} (exothermic effect) and TΔS was 3.6 ± 0.1 kcal $mol^{-1}deg^{-1}$, yielding a Gibbs free energy of -7.5 kcal $mole^{-1}$. Overall, the number of binding sites was less than unity (about 0.7 per mol of sites), which may reflect some endogenous substrate bound to a fraction of the purified protein (see below) and/or an overestimation of the protein concentration.

Substrate specificity of OpuAC

Previously studied glycine betaine transport systems facilitate the accumulation of a wide variety of substrates [22,23]. Binding of a number of these related substrates to OpuAC was determined by intrinsic protein fluorescence measurements as shown in Fig. 3. Proline betaine did not elicit significant fluorescence changes, but a K_I of 41±8 µM could be estimated from a (competitive) inhibition experiment by measuring glycine betaine-induced fluorescence changes. The affinity for proline was low and a dissociation constant of 132±20 mM was estimated. The binding curve for carnitine did not reach a stable end-level, and affinity constants were thus estimated as inhibition constants using glycine betaine as reporter (Fig. 3); the data are summarized in Table 2. Measurements in OpuA-containing proteoliposomes showed transport against a large concentration gradient of glycine betaine but not of carnitine (Fig. 4). Proline slowly entered the proteoliposomes, but this effect was independent of the presence/absence of OpuA. In fact, calculations of the concentration of proline inside the proteoliposomes after 30 min (based on the assumption that the volume of the proteoliposomes is 1 µl/mg of lipid) showed no accumulation of the proline, indicating a diffusion-based process. Overall, the data indicate that OpuAC and thus OpuA is highly specific for glycine betaine and the related proline betaine with only a low affinity for other substrates.

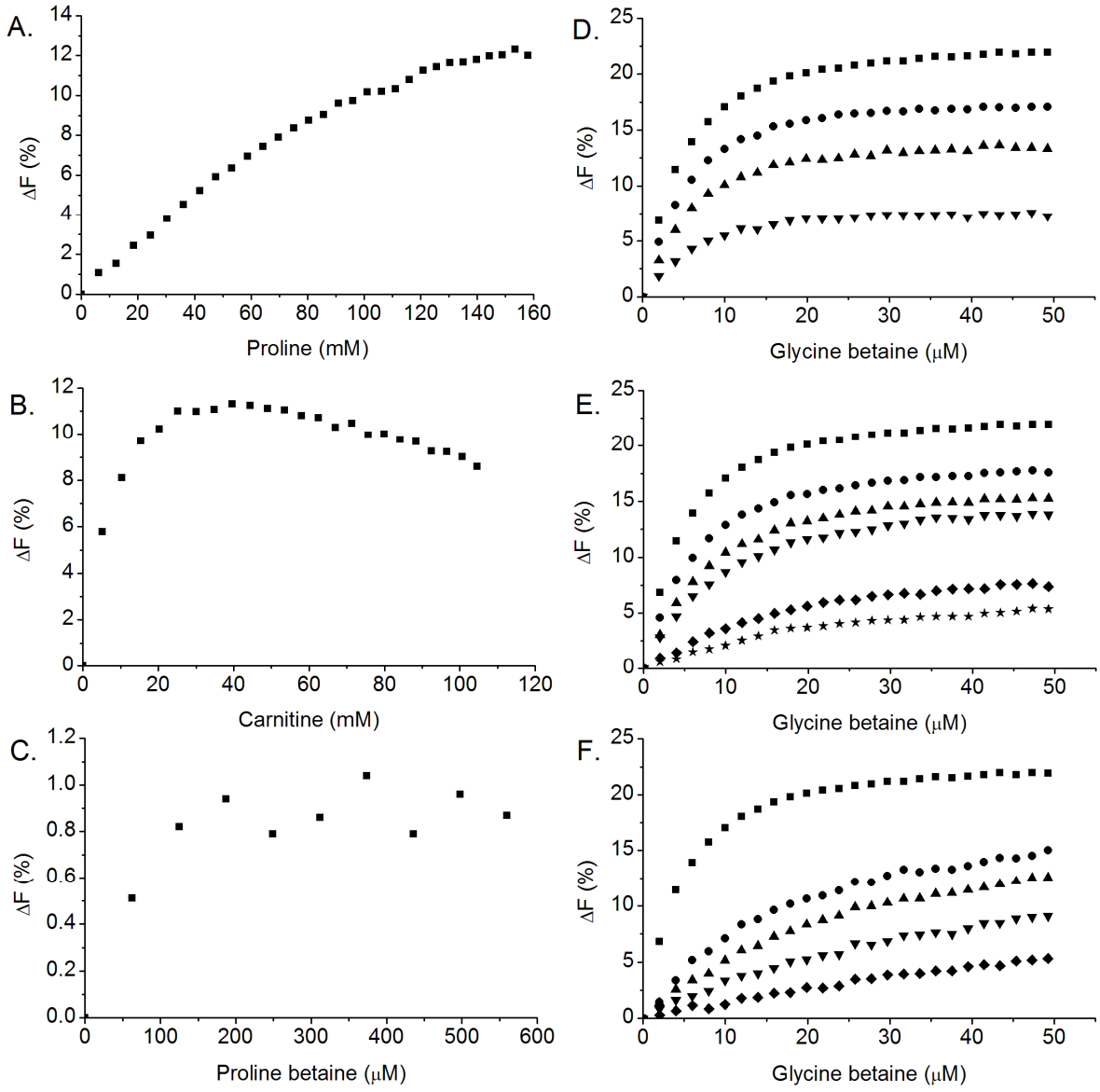


Figure 3. Substrate binding specificity of purified OpuAC. Binding was measured using intrinsic protein fluorescence at an OpuAC concentration of ~0.5 μM. The following substrates were tested: proline (panel A), carnitine (panel B), proline betaine (panel C) and glycine betaine (Panel D-F). In the panels D-F, OpuAC was pre-incubated with different concentrations of proline (panel D), carnitine (panel E) or proline betaine (Panel F) as inhibitor, prior to titration with glycine betaine. In panel D-F, the concentrations of inhibitor are: no inhibitor (■), 25 mM proline/carnitine or 0.125 mM proline betaine (●), 50 mM proline/carnitine or 0.25 mM proline betaine (▲), 100 mM proline/carnitine or 0.5 mM proline betaine (▼), 250 mM carnitine or 1 mM proline betaine (◆) and 500 mM carnitine (★). Measurements in panel D-F were fitted with equation 1 (using a K_D of 4 μM for glycine betaine) and the inhibitor concentrations indicated. The inhibition experiments with proline and carnitine show a decrease in the ΔF_{max} at increased concentrations of inhibitor, which is caused by the high concentrations of inhibitor (see panel A and B). The fluorescence, however, reached the same end level in all measurements, which is indicative for specific and competitive binding.

Table 2. Binding of various ligands to purified OpuAC using intrinsic protein fluorescence. Dissociation constants (K_D) and inhibition constants (K_I) of various ligands bound to OpuAC measured either directly (K_D) or via inhibition of glycine betaine binding (K_I), using intrinsic protein fluorescence. The dissociation constants were averaged from at least duplicate measurements, fitted independently to obtain the dissociation constants (for K_D); for the K_I-values, inhibition was measured over a range of inhibitor concentrations and the K_I values were averaged and the errors indicate the range.

Ligand	K_I	K_D
Proline betaine	41 ± 8 µM	
Proline	198 ± 55 mM	132 ± 20 mM
Carnitine	65 ± 11 mM	

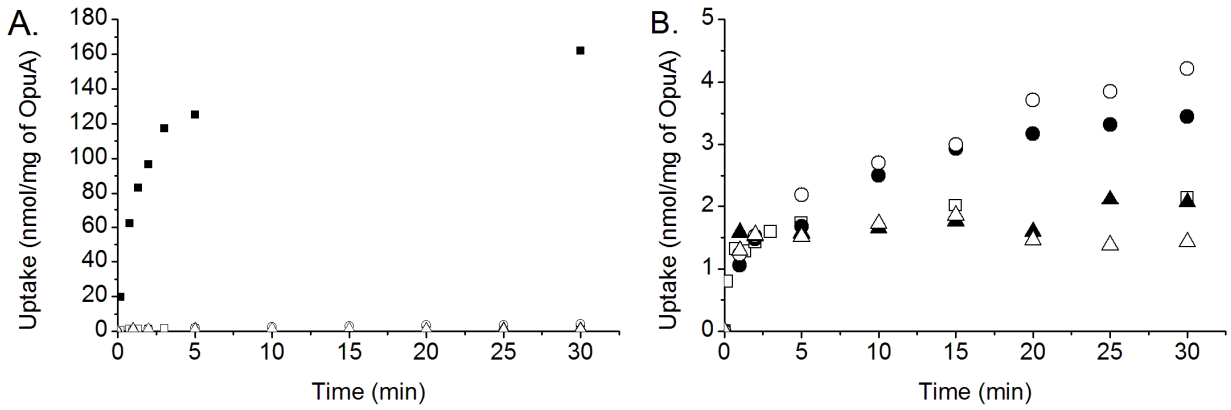


Figure 4. Transport activity of OpuA reconstituted in proteoliposomes composed of 38 mol% DOPG, 50 mol% DOPE and 12 mol% DOPC with the ATP-regeneration system enclosed inside the proteoliposomes. The following ^{14}C-labeled substrates were tested: glycine betaine (■), proline (●) and L-carnitine (▲). Transport was measured using a filter-based assay at a OpuA concentration of 60 µg/ml and with 51 µM (data shown) or 5 mM (not shown) of substrate. The data with 5 mM substrate are not shown but the results were similar as presented in the graphs with 51 µM: glycine betaine was transported against a concentration gradient; carnitine was not taken up, neither in the proteoliposmes nor in the empty liposomes; and proline entered the vesicles slowly but at an equal rate in the proteoliposomes and the liposomes devoid of OpuA. The data thus indicate that only the uptake of glycine betaine is facilitated by OpuA; proline and L-carnitine are bound by the SBDs of OpuA (see Fig. 3) but these substrates are not significantly transported.

3D Structure of OpuAC

The complete OpuA transporter consists of two OpuABC subunits, and thus contains two OpuAC domains. To investigate the possibility that the two substrate-binding domains interact, we determined the oligomeric state of isolated OpuAC by static light scattering coupled to size-exclusion chromatography (SEC-MALLS). OpuAC (Fig. 5) was found to have a molecular mass of 30.3 kDa, irrespective of the presence or absence of 1 mM glycine betaine, which matches a monomeric state of the protein, showing that the isolated domains do not form stable dimers or higher oligomers.

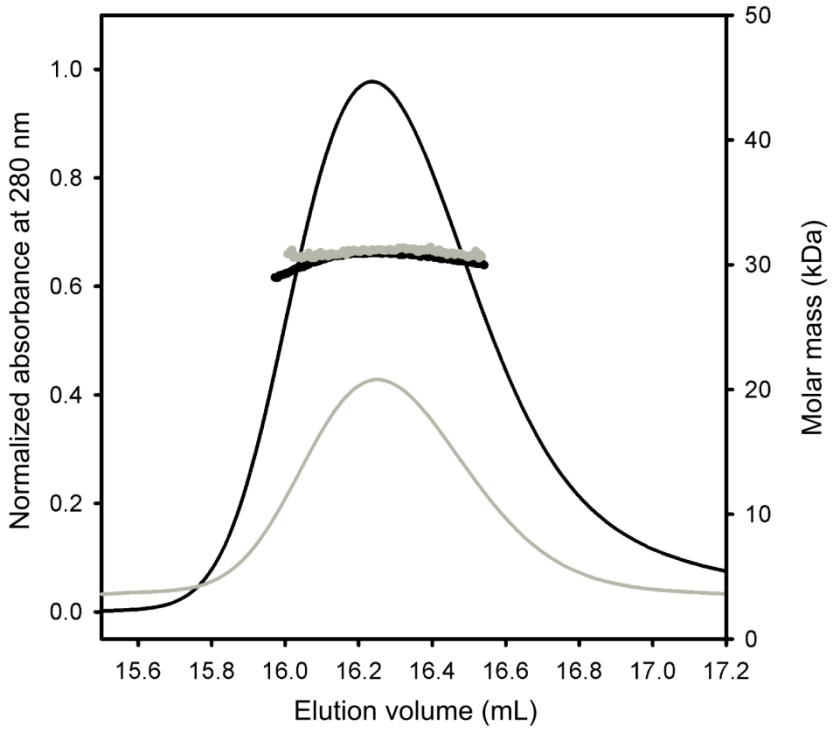


Figure 5. Static light-scattering analysis of purified OpuAC. Purified OpuAC, in the absence or presence of 1 mM glycine betaine, was run on a gel filtration column, coupled to detectors for UV absorbance, refractive index and light scattering. The molecular mass was calculated throughout the elution peaks.

OpuAC was crystallized in the absence and presence of 1 mM glycine betaine, yielding an open conformation (space group $P4_12_12$, diffracting to 1.9 Å resolution) and closed-liganded conformation (space group H32, diffracting to 2.2 Å resolution), respectively (Table 3). The structures were solved by molecular replacement using the individual domains of the *B. subtilis* OpuAC structure as search models (Fig. 6).

The structures of OpuAC showed a fold typical of SBD's, having two α/β domains that enclose a ligand-binding site when in contact with each other (Fig. 6A). OpuAC belongs to Class II of the proteins in the SBD family [24], with two connecting segments forming the hinge, of which the central residues are Lys404 and Ser508. In the open conformation, the α/β domains had moved by 25° relative to one another when compared to the closed-liganded state. This falls in the normal range of opening of SBDs, which in typical class I and II SBDs range from 14° in the binding protein of the leucine transporter [25] to 60° in the binding protein of the leucine/isoleucine/valine transporter, both from *E.coli* [26]. During refinement, electron density accounting for bound ligand became visible in the ligand-binding site of the closed conformation structure. Glycine betaine in the closed-liganded conformation was tightly bound, with B-factors in the same range as for the rest of the protein. In the open conformation structure, electron density was observed in the binding site, although it could not be unambiguously assigned to glycine betaine.

Table 3. Data collection and refinement statistics.

Data collection	**Open structure**	**Closed-liganded structure**
Space group	$P4_12_12$	H32
Cell dimensions	a=b=68.2 Å, c=109.1 Å	a=b=111.7 Å, c=151.7 Å
Wavelength (Å)	0.872	0.872
Unique reflections	24,571	16,390
Completeness (%)	99.7 (98.9)	97.4 (96.1)
Resolution range (Å)	45-1.9	46-2.3
R_{sym}	0.13 (0.45)	0.11 (0.54)
I/σ (I)	15.6 (3.9)	14.4 (4.5)
Average multiplicity	6.0 (6.0)	7.8 (7.3)
Refinement		
Resolution range	29-1.9	40-2.3
Number of reflections	24,526	15,646
R_{work}/R_{free}	0.17/0.21	0.16/0.20
No. atoms		
Protein	2064	2013
Water	285	140
Average *B*-factors ($Å^2$)		
Protein	14.2	18.1
Water	21.6	25.7
R.m.s. deviations		
Bond lengths (Å)	0.005	0.008
Bond angles (°)	0.881	1.086
Ramachandran plot outliers (n, %)	1 (0.4%)	0 (0%)

Ligand binding site

The superposition of the binding-site residues in the open and closed conformation is shown in Fig. 6B. In the closed-liganded conformation, the full binding site has been formed. Compared to the open state, a third tryptophan (W484) has moved in and contributes to the coordination of glycine betaine. The formed Trp-prism, previously described for OpuAC from *B. subtilis* [15], is well suited to coordinate the quaternary ammonium group of glycine betaine via cation-π interactions and van der Waals forces. In the closed-liganded conformation, the carboxylate group of glycine betaine is within optimal distance to form a hydrogen bond to H392, as well as to the backbone of G437 and V438; other residues that move in during the closure of the protein (Fig. 6B, orange).

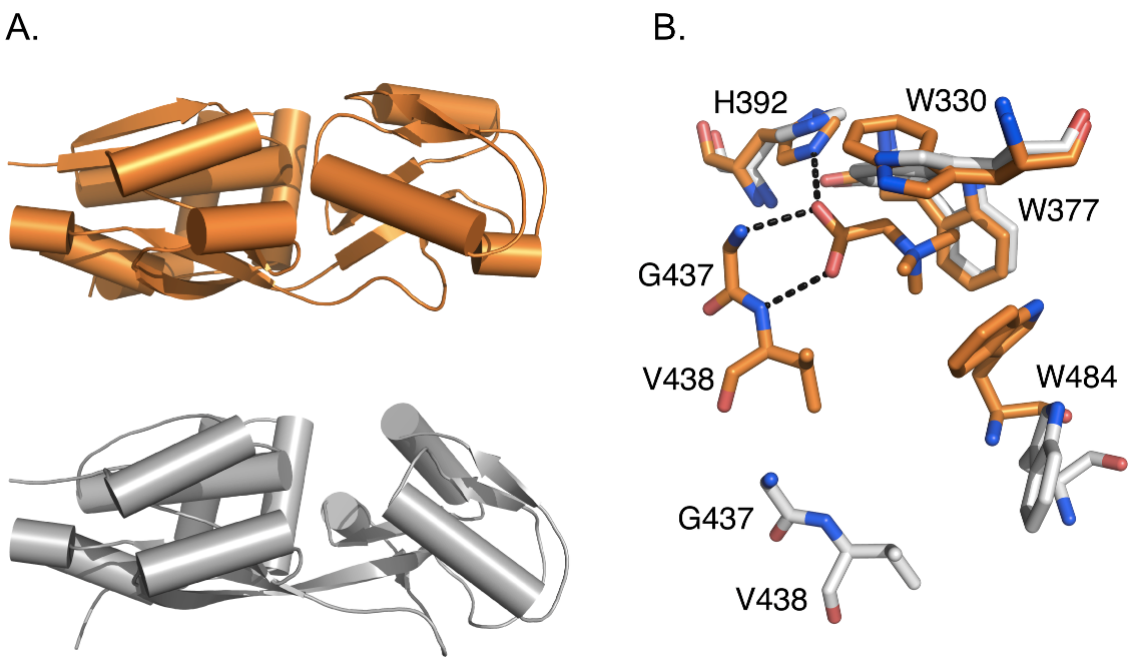


Figure 6. X-ray crystallography structure of OpuAC and its binding site. Panel A shows the structures of OpuAC in its closed (orange) and open (gray) conformations, highlighting the opening of the protein. Panel B, superimposition of the binding pockets for the open and closed-liganded structures of OpuAC. Upon closure of the protein a complete Trp-prism is formed for coordination of the quaternary ammonium moiety of glycine betaine. In addition, hydrogen bonds are formed between the carboxylate of glycine betaine and H392, G437 and V438, thereby stabilizing the closed conformation.

ATP binding to purified OpuA

ATP binding to purified OpuA, OpuAΔSBD and OpuAΔCBS was measured with the fluorescent ATP analog TNP-ATP. TNP-ATP binding to the nucleotide-binding domain gives rise to an increase in the fluorescence together with a small blue shift. The measured fluorescence was corrected for TNP-ATP fluorescence in the buffer. The TNP-ATP binding at 12.5 and 50 mM KP_i pH7 showed no significant difference for the dissociation constants (0.8±0.1 µM and 1.1±0.3 µM, respectively, Fig. 7A-B). At the highest salt concentrations (250 mM KPi, pH7), two different effects were observed. The high salt concentration yielded a higher TNP-ATP fluorescence (Fig. 7A), but also the affinity decreased yielding a dissociation constant of 2.4±0.3 µM (Fig. 7B). The increase in the dissociation constant is small but significant. TNP-ATP binding to truncated versions of OpuA (OpuAΔSBD, which lacks the substrate binding domain and OpuAΔCBS, which lack the osmosensing CBS domains) was measured at 50 mM KP_i, pH 7. The dissociation constant of TNP-ATP binding to OpuAΔCBS was 1.4 ± 0.1 µM, which is comparable to that of the wildtype OpuA protein. On the contrary, deletion of the SBDs significantly lowered the affinity, yielding a dissociation constant of 5.7 ± 0.5 µM (Fig. 7C). The presence of 1 mM glycine betaine did not affect the binding of TNP-ATP to OpuA, OpuAΔSBD and OpuAΔCBS (data not shown). The data are summarized in Table 4. Overall, the data indicate that the previously described ionic control of OpuA activity

[1,2] is not at the level of substrate- or nucleotide-binding, leaving regulation of translocation as the most plausible scenario.

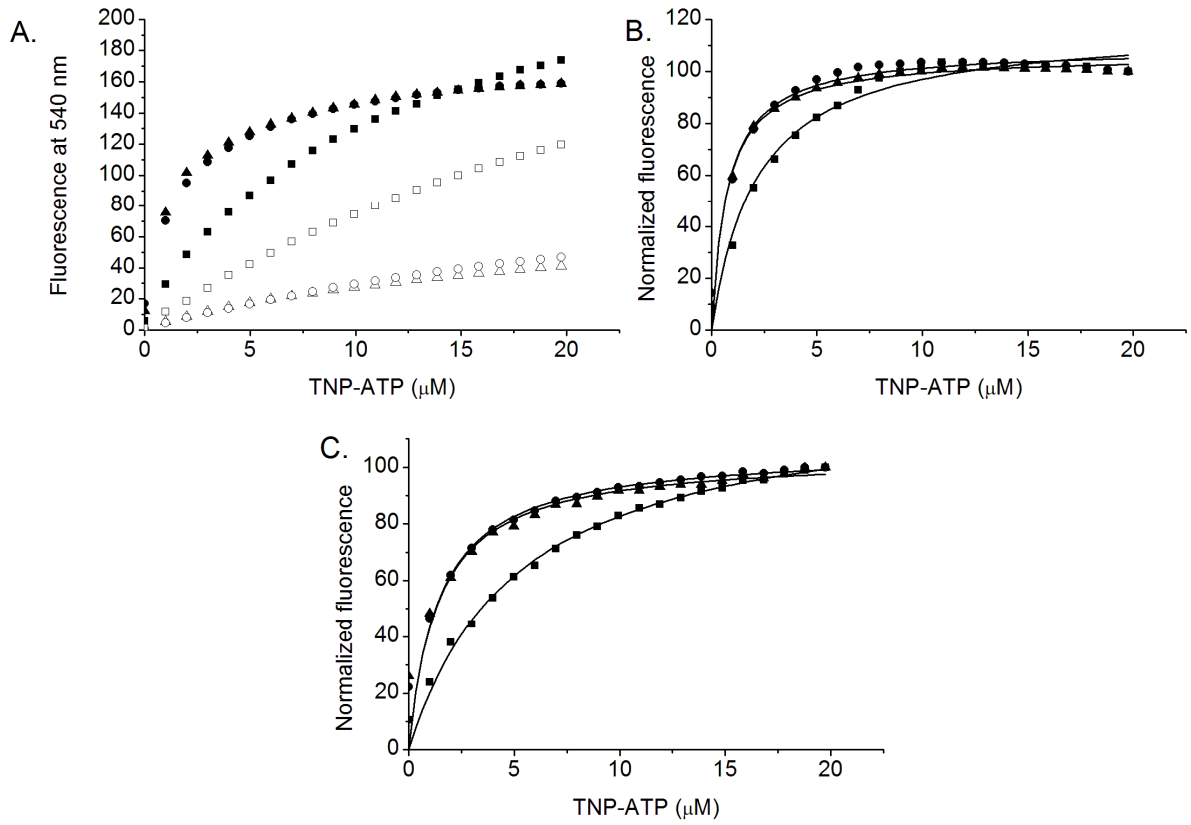


Figure 7. TNP-ATP binding to purified OpuA and derivates. Fluoresence was measured via fluorescence changes upon TNP-ATP binding to the protein. TNP-ATP fluorescence in buffer and TNP-ATP binding to OpuA (with 0.5 µM purified OpuA) was measured at different salt concentrations (12.5-250 mM KP_i pH 7.0) (panel A). TNP-ATP binding to OpuA was corrected for background fluorescence in buffer and normalized to the endlevel (panel B). In panel A and B, measurements were done at 250 mM KP_i (■,□), 50 mM KP_i (●,❍) and 12.5 mM KP_i (▲,△). Closed symbols were used for TNP-ATP binding to OpuA and open symbols were used for the TNP-ATP fluorescence in buffer, the data fits were shown as black lines. In panel C, TNP-ATP binding to OpuA (▲), OpuAΔCBS (●) and OpuAΔSBD (■) was measured at 50 mM KP_i pH 7.

Table 4. TNP-ATP binding to purified OpuA, OpuAΔSBD and OpuAΔCBS. The dissociation constants of the binding were averaged from at least two independent measurements and the errors indicate the range.

KP_i, pH7.0 (mM)	OpuA K_D (µM)	OpuAΔCBS K_D (µM)	OpuAΔSBD K_D (µM)
12.5	0.81 ± 0.09		
50	1.09 ± 0.33	1.36 ± 0.05	5.66 ± 0.52
250	2.41 ± 0.26		

Discussion

The osmoregulatory ABC transporter OpuA accumulates glycine betaine in response to osmotic stress. Although ionic strength is the signal of the osmotic sensing mechanism, we now show that glycine betaine binding to the SBDs of OpuA is independent of the salt concentration in the medium. This was established in binding studies using the full size OpuA protein in a detergent-solubilized state as well as the isolated SBD domains expressed as soluble protein (OpuAC). Some glycine betaine transport systems have rather broad substrate specificity, *e.g.* ProP from *E. coli,* which transports among others proline, glycine betaine and ectoine [23], and OpuC from *B. subtilis* which takes glycine betaine, crotonobetaine and carnitine together with more substrates [22]. Others, like BetP from *C. glutamicum*, are more specific to glycine betaine [27]. OpuA falls into the category of highly specific systems, as the dissociation constant for proline and carnitine is at least four orders of magnitude higher than that of glycine betaine.

Structural basis for substrate-binding

The dissociation constant for glycine betaine binding (K_D ~ 4 µM) to OpuAC (and thus OpuA) from *L. lactis* is about 4-fold lower than the corresponding value for the OpuAC homolog from *B. subtilis* (K_D ~17 µM, [15]); similar differences are observed for proline betaine (K_D ~41 µM and ~295 µM, respectively [15]). At the level of primary sequence, OpuAC from *L. lactis* and *B. subtilis* are 44% identical. To find out what caused these differences, OpuAC was crystallized and its structure determined. Superimposing the Cα-backbone structures of OpuAC from *L. lactis* and *B. subtilis* yields an r.m.s.d of 0.98 Å. If one only superimposes the residues involved in the ligand binding, the r.m.s.d is 0.28 Å, implying that the binding sites are virtually identical. Since the binding sites do not differ, the explanation for the higher affinities in OpuAC from *L. lactis* must lie elsewhere in the protein. It has previously been shown for maltose-binding protein that mutating a residue which allosterically affects the equilibrium between the apo-state and the closed-liganded conformation can change the K_D for ligand binding by more than 2-orders of magnitude, while leaving the binding site intact [28]. The differences in substrate-binding activity between OpuAC from *L. lactis* and *B. subtilis* may have a similar basis and be caused by a lower flexibility in the *L. lactis* protein, which would imply a lower entropic barrier for domain closure and a decrease in k_{off} (and thus K_d) for the ligands.

Electron density was observed in the ligand-binding site of the open conformation structure. This density could not unambiguously be determined, but likely represents endogenous ligand that was co-purified with the protein, as ITC and fluorescence titration experiments showed that the amount of substrate bound per mol OpuAC was below 1. On the basis of modeling and refinement of the ambiguous density as glycine betaine, the initial coordination of the quaternary ammonium moiety would be performed by W377.

Several structures of proteins that bind quaternary ammonium compounds like glycine betaine have now been elucidated (e.g. BetP (PDB code 2WIT), AChBP (PDB code 1UV6), ChoX (PDB code 2REG), ProX from *E. coli* (PDB code 1R9L), and ProX from *A. fulgides* (PDB code 1SW2) [29-33]. ChoX from *S. meliloti* and ProX from *E. coli* are both SBPs with a similar fold as OpuAC, whereas BetP from *C. glutamicum* is a trimeric Na^+-coupled symporter for glycine betaine. AChBP is a soluble homolog of the ligand-binding domain of nicotinic acetylcholine receptors, serving as a structural model for the pharmaceutically important family of pentameric ligand-gated ion channels [32,34,35]. Even though the overall structures of BetP, AChBP and OpuAC are unrelated, these proteins all share a common binding motif, with aromatic residues forming a box around the ligand and interacting via cation-π interactions. Likely this aromatic box is the optimal solution for the binding of the quaternary ammonium moiety. It is tempting to speculate that these systems have converged during evolution, very much like the independently evolved catalytic triad (Ser-His-Asp) in proteases [36].

Glycine betaine binding to OpuA is dependent on the environment of the protein. In a lipidic environment the dissociation constant is around 0.5 µM [14]), but in the DDM-solubilized state the dissociation constant is increased around 10-fold. It has previously been established that the lipidic environment is important for regulation of the OpuA transport activity [1]. Here we show that the membrane environment affects ligand binding as well but the effect is not related to the ionic sensing mechanism described previously [1].

Nucleotide-binding to OpuA

Biochemical studies have established a salt dependence for the nucleotide-binding subunit (OpuAA) of the glycine betaine transporter from *B. subtilis*. *B.subtilis* OpuAA undergoes a dynamic monomer–dimer equilibration forming dimers in the nucleotide-free and ATP-bound states, but converts to a monomeric state in the ADP-bound form. At high salt concentrations the interconversion between these two states is diminished [37]. Nucleotide (TNP-ATP) binding to *B. subtilis* OpuAA, on the other hand, showed hardly any ionic strength dependence (in the range of 150-1000 mM NaCl [37]), similar to the binding we measured with the full length OpuA from *L. lactis* (in the range of 12.5 to 250 mM KP_i). The absence of the substrate-binding domain had a negative effect on the TNP-ATP binding, which was reflected by an increased dissociation constant for TNP-ATP binding to OpuAΔSBD. This effect was not induced by glycine betaine. Thus, irrespective of the binding of glycine betaine, the presence of the substrate-binding domain seems to have a positive effect on the TNP-ATP binding, which must be signaled via the membrane domain. The comparable affinity of TNP-ATP for wildtype OpuA and OpuAΔCBS indicates that the CBS domains do not influence nucleotide binding. The ability of CBS modules to bind nucleotides was published by Scott et al. [38], but attempts to measure ^{32}γ-ATP binding to the CBS domains of OpuA did not yield

detectable signals (unpublished results). Together, these experiments indicate that CBS modules of OpuA do not participate in the binding of nucleotides.

Conclusions

Substrate binding to OpuA was not affected by ionic strength, implying that ionic control of transport is at a later step. Also, the cooperativity observed in transport [14,39] is not observed when binding of glycine betaine is probed. Overall, the data underline the importance of the lipidic membrane for the osmosensing and regulation of OpuA. The crystal structures of OpuAC show a classic Trp-prism binding pocket, forming the basis for the high affinity binding of glycine betaine, and reveal that the differences in ligand affinities between OpuAC from *L. lactis* and *B. subtilis* must lie outside of the ligand binding site.

Material and Methods

Bacterial strains and growth conditions

Plasmids for expression of OpuA or derivatives were propagated in *Lactococcus lactis* strain NZ9000. The following OpuA mutants were used: OpuA□CBS [1], and OpuAΔSBD, OpuAC [14]. The topologies of the different constructs are depicted schematically in Fig. 1A. The strains were cultivated semi-anaerobically in 2% (w/v) Gistex LS (Strik BV, Eemnes, NL), 65 mM potassium phosphate (KPi) pH 6.5, 1.0 % (w/v) glucose and 5 µg/ml chloramphenicol at 30 °C. For the isolation of the membrane vesicles, cells were grown in a 2 liter pH-regulated bioreactor to an OD_{600} of 2, after which transcription from the *nisA* promoter was initiated by the addition of 0.1 % (v/v) culture supernatant from the nisin A producing strain NZ9700. The cells were harvested and stored at -80°C for preparation of membrane vesicles.

Preparation membrane vesicles and protein purification

Cell disruption was performed by two passages through a cell disruptor (Constant Systems Ltd) at 39 kPsi in the presence of 1 mM $MgSO_4$, 1 mM PMSF, 100 µg/mL DNase and 100 µg/mL RNase. The cell debris was removed from the solution by centrifugation (15 min 11,814 x*g* at 4 °C) and the membrane vesicles were collected by ultracentrifugation (60 min 185,000 x*g* at 4°C). The pellet fraction (containing the membrane vesicles) was resuspended in 50 mM KP_i, pH 7, 20% glycerol (Buffer A) to a concentration of ca. 20 mg/mL total protein. In case of preparation of the soluble OpuAC, the cell debris and membrane vesicles were spun down by ultracentrifugation in the presence of 200 mM KCl to prevent unspecific binding of OpuAC to the membrane vesicles.

Purification of OpuA

Membrane vesicles were collected by centrifugation (20 min 267,008 x*g* at 4°C) and resuspended in buffer A plus 200 mM KCl to a final concentration of 5 mg/mL total

protein. The membrane vesicles were solubilized in 0.5% DDM for 30 min on ice. The insoluble fraction was removed by ultracentrifugation (20 min 267,008 x*g* at 4°C), after which the solubilized material was diluted five-fold in buffer A, supplemented with 200 mM KCl and 15 mM imidazole pH 8.0 to reduce the detergent concentration. The solubilisate was incubated with pre-equilibrated Nickel-Sepharose for 1-2 hours at 4 °C under rotation, after which the resin (bed volume of 0.5 ml) was poured into a 10 mL column. The Nickel-Sepharose resin was washed with 10 mL of Buffer A containing 200 mM KCl, 15 mM imidazole and 0.05% DDM, after which the protein was eluted with Buffer A supplemented with 200 mM KCl ,200 mM imidazole and 0.05% DDM. OpuAΔCBS was purified using similar conditions, except that the imidazole concentration was increased to 50 mM imidazole for incubation and washing and 500 mM imidazole for elution of the protein.

Purification of OpuAC

The cell lysate from the disrupted cells after ultracentrifugation (supplemented with 50 mM imidazole) was mixed immediately with pre-equilibrated Nickel-Sepharose. Washing and elution was done in 50 mM KP_i pH 7, supplemented with 50 and 500 mM imidazole, respectively. OpuAC was purified further on a Superdex 200 10/300 Gl column in either 20 mM Na-MES pH 6.0, 150 mM NaCl (for crystallization purposes) or in 150 mM KP_i pH 7.0 (for biochemical characterization). Fractions containing OpuAC were pooled and concentrated, using a Vivaspin column with a cut-off of 10 kDa.

Fluorescence measurements

Fluorescence spectra were obtained with a Fluorolog®-3 (Jobin Yvon) spectrofluorometer. A quartz-cuvette containing 800 µL of protein sample was incubated for 10 minutes at 25 °C under constant stirring before stepwise addition of the substrate (or buffer as control), using a Hamilton syringe pump (Harvard apparatus). Samples were incubated for 5 seconds before the fluorescence signal was collected for 20 seconds to obtain an averaged value at each substrate concentration.

For fluorescence titration experiments, around 0.5 µM of purified OpuAC (in 150 mM KP_i, pH 7) or OpuA (in 50mM KP_i pH 7, 20% glycerol, 0.05% DDM) was used, and solutions of glycine betaine and TNP-ATP were added in steps of 0.5 µL. For the intrinsic protein fluorescence measurements, the excitation and emission wavelengths were 295 and 360 nm (and slit widths of 1 and 5 nm), respectively. For the measurements of TNP-ATP binding to OpuA, the excitation and emission wavelengths were 409 and 540 nm (slit widths of 1 and 5 nm), respectively. To determine the salt dependence of ligand-binding to purified OpuA, the buffer was exchanged to various KP_i concentrations (12.5-250 mM), pH 7.0, 20% glycerol and 0.05% DDM, using a NAP10 column (GE Healthcare). Corrections for background fluorescence changes were made by titrations with buffer (intrinsic protein fluorescence), or titrations to a reference sample without protein (TNP-ATP binding).

Data analysis

Fluorescence titrations were analyzed essentially as described by Lanfermeijer [17]. Curve fitting was performed in Origin (OriginLab). The dissociation constants for competitive binding were calculated using equation 1,

$K_{D,app} = K_d(1+s[I]/K_i)$ *(equation 1)*

in which $K_{D,app}$ is the apparent K_D after fitting, K_D is the affinity constant for glycine betaine binding, [I] is the concentration of the competitive ligand, K_I corresponds to the K_D of the competitive ligand.

Isothermal titration calorimetry (ITC)

Glycine betaine binding to OpuAC was measured by microcalorimetry on a ITC_{200} calorimeter (MicroCal) at 25 °C. 200 µL of OpuAC (concentrated to 50 µM in 150 mM KP_i, pH 7) was added to the cell. To determine the binding constant, glycine betaine (500 µM stock solution in the same buffer as the protein) was added stepwise. Typically, 20 injections of 2 µL volume were made with intervals of 120 seconds between each addition. The first titration in each experiment was 0.5 µL of glycine betaine instead of 2 µL, which was subsequently deleted in the data analysis; data were analyzed using the MicroCal software provided [40].

Glycine betaine binding by filter-based assay (protein precipitation)

Binding of radio-labeled glycine betaine was also measured by the precipitation method [18]. This method is based on the principle that upon salting-out of the protein (by ammonium sulfate), the substrate remains trapped in the ligand-binding site. For these measurements, purified OpuA in the presence of 0.05% (w/v) *n*-dodecyl-β-maltopyranoside (DDM) was mixed such that the protein (at a final concentration of 100 µg/mL) was in a buffer composed of 10, 50 or 250 mM KP_i pH7.0, 20% glycerol and 0.05% DDM. The binding assays were carried in a volume of 100 µL. After 2 min of pre-incubation at 30 °C, [^{3}H]-glycine betaine was added to the assay mixture at the desired concentration (ranging from 0.1 µM to 42.5 µM) and the binding reaction was quenched after 2 min by dilution of the sample into 2 ml ice-cold 50 % (w/v) ammonium sulfate solution. The mixture was filtered rapidly through 0.45 µm pore-size cellulose nitrate filters. The filters were washed twice with 2 ml ammonium sulfate solution. Subsequently, the filters were placed in an open plastic vial and dried under heating in an oven at 37 °C for 2 hours. The radioactivity on the filters was measured via liquid scintillation counting, using emulsifier plus scintillation liquid (Perkin Elmer). Measurements were corrected for background signals and the average of the corrected data was fitted as described under fluorescence measurements. The specific activity of ^{14}C-labeled glycine betaine was high enough for the transport assays (see below) but not for the ligand binding studies. We thus used ^{3}H-glycine betaine in the binding studies. Radio-labeled [^{3}H]-glycine betaine was prepared via a conversion of [^{3}H]-choline chloride

(Amersham, specific activity: $2.01*10^6$ MBq/mmol) to glycine betaine as described by Boch et al. [41].

Substrate transport

The transport activity of OpuA was measured using [^{14}C]-glycine betaine (converted from [^{14}C]-choline chloride (Amersham, specific activity: $2.07*10^3$ MBq/mmol) as described by Boch et al. [41]), [^{14}C]-L-proline (Amersham, specific activity: $8.58*10^3$ MBq/mmol) and [^{14}C]-L-carnitine (GE Healthcare, specific activity: $2.11*10^3$ MBq/mmol) with OpuA reconstituted in proteoliposomes containing 38 mol% dioleoyl-phosphatidylglycerol, 50 mol% dioleoyl-phosphatidylethanolamine and 12 mol% dioleoyl-phosphatidylcholine. Proteoliposomes were prepared as described previously by Geertsma et al. [42] with an ATP regenerating system present and using 200 mM KP_i, pH 7, as external medium. The transport activity was measured using a filter-based assay as described by Geertsma et al. [42] in a time dependent way (up to 30 min), using a fixed concentration of 50 µM or 5 mM of substrate.

Static light scattering

An aliquot of 200 µl of OpuAC (0.2-0.4 mg/ml) was run at a flow rate of 0.5 ml/min on a Superdex 200 10/300GL gel filtration column (GE Healthcare) in 150 mM KP_i pH 7.0 (+/- 1 mM GB) using an Agilent 1200 series isocratic pump at room temperature. Detectors were used for absorbance at 280nm (Agilent), static light-scattering (miniDawn TREOS Wyatt) and differential refractive index (Optilab Rex Wyatt). For data analysis, the ASTRA software package version 5.3.2.10 was used (Wyatt), with a value for the refractive index increment (dn/dc) protein of 0.187 ml/mg [43,44].

Crystallization and structure determination

OpuAC, in 20 mM Na-MES, pH 6.0 and 15 mM NaCl, was concentrated to 9 mg/mL. Crystals of OpuAC were grown by vapor diffusion in hanging drops. Crystallization conditions that yielded the open-liganded conformation consisted of 1 µL protein (9 mg/mL OpuAC) and 1 µL reservoir solution (0.2 M sodium iodide, 0.1 M Bis-Tris propane pH 8.5, 20% w/v PEG 3350). Crystals yielding the closed-liganded conformation were grown with 1 µL protein (9 mg/mL OpuAC plus 1 mM glycine betaine) and 1 µL reservoir solution (0.2 M NaCl, 0.1 M Na-Hepes, pH 7.0, 20% PEG 6000). Crystals of the open conformation were obtained after 21 days of incubation at 18 °C, while for the closed-liganded conformation crystals were obtained after 4 days of incubation at 18 °C. Crystals of the open and closed–liganded conformation were soaked in mother liquor supplemented with 12% glycerol for 30 s and 42% PEG 6000 for 30 s, respectively, and then flash cooled in liquid nitrogen. Data was collected to 1.9 Å and 2.3 Å resolution for the open and closed-liganded conformation, respectively, on beamline ID23-2 at the ESRF, Grenoble. Data processing, reduction and scaling were carried out using the program XDS [45]. The structures were solved by molecular replacement with the program Phaser [46], using the *B. subtilis* OpuAC structure (PDB code: 2b4l) as a search

model. A few cycles of refinement using Refmac5 [47] and Phenix.refine [48], interspersed with manual model building using Coot [49], were necessary to complete the model. Water molecules were placed automatically in Fo-Fc Fourier difference maps at a 3 σ-cutoff level, and validated to ensure correct coordination geometries using Coot. Relevant statistics of the data collection and model refinement are given in Table 3.

Accession codes

The coordinates have been deposited in the Proteins Data Bank with accession codes 3L6G and 3L6H.

Acknowledgements

The authors are grateful to Prof. Lutz Schmitt for kindly supplying proline betaine.

References

[1] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. USA*, **103,** 10624-10629.

[2] Mahmood,N.A., Biemans-Oldehinkel,E., Patzlaff,J.S., Schuurman-Wolters,G.K., & Poolman,B. (2006) Ion specificity and ionic strength dependence of the osmoregulatory ABC transporter OpuA. *J. Biol. Chem*, **281,** 29830-29839.

[3] van der Heide,T. & Poolman,B. (2000) Glycine betaine transport in Lactococcus lactis is osmotically regulated at the level of expression and translocation activity. *J. Bacteriol*, **182,** 203-206.

[4] Patzlaff,J.S., van der Heide,T., & Poolman,B. (2003) The ATP/Substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem*, **278,** 29546-29551.

[5] Mahmood,N.A.B.N., Biemans-Oldehinkel,E., & Poolman,B. (2009) Engineering of Ion Sensing by the Cystathionine beta-Synthase Module of the ABC Transporter OpuA. *J. Biol. Chem*, **284,** 14368-14376.

[6] Chen,C.L. & Beattie,G.A. (2007) Characterization of the osmoprotectant transporter OpuC from Pseudomonas syringae and demonstration that cystathionine-beta-synthase domains are required for its osmoregulatory function. *J. Bacteriol*, **189,** 6901-6912.

[7] Culham,D.E., Romantsov,T., & Wood,J.M. (2008) Roles of K+, H+, H2O, and DeltaPsi in solute transport mediated by major facilitator superfamily members ProP and LacY. *Biochemistry*, **47,** 8176-8185.

[8] Ott,V., Koch,J., Spate,K., Morbach,S., & Kramer,R. (2008) Regulatory properties and interaction of the C- and N-terminal domains of BetP, an osmoregulated betaine transporter from Corynebacterium glutamicum. *Biochemistry*, **47,** 12208-12218.

[9] Schiller,D., Kramer,R., & Morbach,S. (2004) Cation specificity of osmosensing by the betaine carrier BetP of Corynebacterium glutamicum. *Febs Letters*, **563,** 108-112.

[10] Culham,D.E., Hillar,A., Henderson,J., Ly,A., Vernikovska,Y.I., Racher,K.I., Boggs,J.M., & Wood,J.M. (2003) Creation of a fully functional cysteine-less variant of osmosensor and proton-osmoprotectant symporter ProP from

Escherichia coli and its application to assess the transporter's membrane orientation. *Biochemistry*, **42,** 11815-11823.

[11] Romeo,Y., Bouvier,J., & Gutierrez,C. (2007) Osmotic regulation of transcription in Lactococcus lactis: ionic strength-dependent binding of the BusR repressor to the busA promoter. *FEBS Lett*, **581,** 3387-3390.

[12] Monnet,V. (2003) Bacterial oligopeptide-binding proteins. *Cell Mol Life Sci*, **60,** 2100-2114.

[13] Mao,B., Pear,M.R., Mccammon,J.A., & Quiocho,F.A. (1982) Hinge-Bending in L-Arabinose-Binding Protein - the Venus-Flytrap Model. *J. Biol. Chem*, **257,** 1131-1133.

[14] Biemans-Oldehinkel,E. & Poolman,B. (2003) On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. *EMBO J.*, **22,** 5983-5993.

[15] Horn,C., Sohn-Bosser,L., Breed,J., Welte,W., Schmitt,L., & Bremer,E. (2006) Molecular determinants for substrate specificity of the ligand-binding protein OpuAC from Bacillus subtilis for the compatible solutes glycine betaine and proline betaine. *J. Mol. Biol*, **357,** 592-606.

[16] Smits,S.H., Hoing,M., Lecher,J., Jebbar,M., Schmitt,L., & Bremer,E. (2008) The compatible-solute-binding protein OpuAC from Bacillus subtilis: ligand binding, site-directed mutagenesis, and crystallographic studies. *J. Bacteriol*, **190,** 5663-5671.

[17] Lanfermeijer,F.C., Picon,A., Konings,W.N., & Poolman,B. (1999) Kinetics and consequences of binding of nona- and dodecapeptides to the oligopeptide binding protein (OppA) of Lactococcus lactis. *Biochemistry*, **38,** 14440-14450.

[18] Richarme,G. & Kepes,A. (1983) Study of Binding Protein-Ligand Interaction by Ammonium Sulfate-Assisted Adsorption on Cellulose Esters Filters. *Biochim. Biophys. Acta*, **742,** 16-24.

[19] Gerber,S., Comellas-Bigler,M., Goetz,B.A., & Locher,K.P. (2008) Structural basis of trans-inhibition in a molybdate/tungstate ABC transporter. *Science*, **321,** 246-250.

[20] Kadaba,N.S., Kaiser,J.T., Johnson,E., Lee,A., & Rees,D.C. (2008) The high-affinity E. coli methionine ABC transporter: structure and allosteric regulation. *Science*, **321,** 250-253.

[21] Oldham,M.L., Khare,D., Quiocho,F.A., Davidson,A.L., & Chen,J. (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*, **450,** 515-521.

[22] Kappes,R.M. & Bremer,E. (1998) Response of Bacillus subtilis to high osmolarity: uptake of carnitine, crotonobetaine and gamma-butyrobetaine via the ABC transport system OpuC. *Microbiol.*, **144,** 83-90.

[23] MacMillan,S.V., Alexander,D.A., Culham,D.E., Kunte,H.J., Marshall,E.V., Rochon,D., & Wood,J.M. (1999) The ion coupling and organic substrate specificities of osmoregulatory transporter ProP in Escherichia coli. *Biochim. Biophys. Acta*, **1420,** 30-44.

[24] Quiocho,F.A. & Ledvina,P.S. (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: Variation of common themes. *Mol. Microbiol*, **20,** 17-25.

[25] Magnusson,U., Salopek-Sondi,B., Luck,L.A., & Mowbray,S.L. (2004) X-ray structures of the leucine-binding protein illustrate conformational changes and the basis of ligand specificity. *J Biol Chem*, **279,** 8747-8752.

[26] Trakhanov,S., Vyas,N.K., Luecke,H., Kristensen,D.M., Ma,J., & Quiocho,F.A. (2005) Ligand-free and -bound structures of the binding protein (LivJ) of the Escherichia coli ABC leucine/isoleucine/valine transport system: trajectory and dynamics of the interdomain rotation and ligand specificity. *Biochemistry*, **44,** 6597-6608.

[27] Peter,H., Burkovski,A., & Kramer,R. (1996) Isolation, characterization, and expression of the Corynebacterium glutamicum betP gene, encoding the transport system for the compatible solute glycine betaine. *J. Bacteriol*, **178,** 5229-5234.

[28] Marvin,J.S. & Hellinga,H.W. (2001) Manipulation of ligand binding affinity by exploitation of conformational coupling. *Nature Struct. Biol*, **8,** 795-798.

[29] Oswald,C., Smits,S.H.J., Hoing,M., Sohn-Bosser,L., Dupont,L., Le Rudulier,D., Schmitt,L., & Bremer,E. (2008) Crystal Structures of the Choline/Acetylcholine Substrate-binding Protein ChoX from Sinorhizobium meliloti in the Liganded and Unliganded-Closed States. *J. Biol. Chem*, **283,** 32848-32859.

[30] Ressl,S., van Scheltinga,A.C.T., Vonrhein,C., Ott,V., & Ziegler,C. (2009) Molecular basis of transport and regulation in the Na+/betaine symporter BetP. *Nature*, **458,** 47-52.

[31] Schiefner,A., Breed,J., Bosser,L., Kneip,S., Gade,J., Holtmann,G., Diederichs,K., Welte,W., & Bremer,E. (2004) Cation-pi interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli. *J. Biol. Chem*, **279,** 5588-5596.

[32] Celie,P.H.N., Rossum-Fikkert,S.E., van Dijk,W.J., Brejc,K., Smit,A.B., & Sixma,T.K. (2004) Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron*, **41,** 907-914.

[33] Schiefner,A., Holtmann,G., Diederichs,K., Welte,W., & Bremer,E. (2004) Structural basis for the binding of compatible solutes by ProX from the hyperthermophilic archaeon Archaeoglobus fulgidus. *J. Biol. Chem*, **279,** 48270-48281.

[34] Brejc,K., van Dijk,W.J., Klaassen,R.V., Schuurmans,M., van der Oost,J., Smit,A.B., & Sixma,T.K. (2001) Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature*, **411,** 269-276.

[35] Hilf,R.J.C. & Dutzler,R. (2009) Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature*, **457,** 115-118.

[36] Perona,J.J. & Craik,C.S. (1995) Structural Basis of Substrate-Specificity in the Serine Proteases. *Protein Sci*, **4,** 337-360.

[37] Horn,C., Bremer,E., & Schmitt,L. (2003) Nucleotide dependent monomer/dimer equilibrium of OpuAA, the nucleotide-binding protein of the osmotically regulated ABC transporter OpuA from Bacillus subtilis. *J. Mol. Biol*, **334,** 403-419.

[38] Scott,J.W., Hawley,S.A., Green,K.A., Anis,M., Stewart,G., Scullion,G.A., Norman,D.G., & Hardie,D.G. (2004) CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J. Clin. Invest*, **113,** 274-284.

[39] van der Heide,T. & Poolman,B. (2000) Osmoregulated ABC-transport system of Lactococcus lactis senses water stress via changes in the physical state of the membrane. *Proc. Natl. Acad. Sci. USA*, **97,** 7102-7106.
[40] Wiseman,T., Williston,S., Brandts,J.F., & Lin,L.N. (1989) Rapid Measurement of Binding Constants and Heats of Binding Using A New Titration Calorimeter. *Anal. Biochem*, **179,** 131-137.
[41] Boch,J., Kempf,B., & Bremer,E. (1994) Osmoregulation in Bacillus subtilis: synthesis of the osmoprotectant glycine betaine from exogenously provided choline. *J Bacteriol*, **176,** 5364-5371.
[42] Geertsma,E.R., Mahmood,N.A.B.N., Schuurman-Wolters,G.K., & Poolman,B. (2008) Membrane reconstitution of ABC transporters and assays of translocator function. *Nat. Protoc*, **3,** 256-266.
[43] Folta-Stogniew,E. & Williams,K. (1999) Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. *J. Biomol. Tech*, **10,** 51-63.
[44] Slotboom,D.J., Duurkens,R.H., Olieman,K., & Erkens,G.B. (2008) Static light scattering to characterize membrane proteins in detergent solution. *Methods*, **46,** 73-82.
[45] Kabsch,W. (1993) Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Crystallogr*, **26,** 795-800.
[46] Mccoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C., & Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr*, **40,** 658-674.
[47] Murshudov,G.N., Vagin,A.A., & Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **53,** 240-255.
[48] Adams,P.D., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., Mccoy,A.J., Moriarty,N.W., Read,R.J., Sacchettini,J.C., Sauter,N.K., & Terwilliger,T.C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.*, **58,** 1948-1954.
[49] Emsley,P. & Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr.*, **60,** 2126-2132.

Chapter 5 The response of *Lactococcus lactis* to membrane protein production

Justina.C.Wolters[#], Ravi K. R. Marreddy[#], Joao P. C. Pinto[#], Eric R. Geertsma, Fabrizia Fusetti, Hjalmar P. Permentier, Jan Kok and Bert Poolman

[#] These authors contributed equally to the work

Abstract

The recombinant expression of membrane proteins in functional form and amounts high enough for structural and functional studies is often problematic. To engineer cells towards efficient membrane protein production, it is important to characterize the stress response to the expression. We employed a combined proteomics and transcriptomics approach to probe the cellular responses upon the synthesis of membrane proteins in *Lactococcus lactis*. The overproduction of membrane protein resulted in severe growth defects, presumably due to a diversion of nutrient towards the synthesis of recombinant proteins and/or too limited capacity to import amino acids. In addition membrane protein overproduction invoked a general stress response (upregulation of various chaperones and proteases), which is probably due to accumulation of misfolded protein. Most remarkably, a two component (CesSR)-mediated cell envelope stress response was observed, which targeted several proteins for subsequent engineering and better understanding of the bottlenecks in membrane protein production.

Introduction

Membrane proteins are involved in many essential cellular processes such as transport of nutrients, sensing of environmental changes, energy transduction and scaffolding of cell structure. Due to their important roles in various diseases these proteins are clinically important as potential drug targets. To date, two thirds of all available pharmaceutical drugs target membrane proteins [1]. Even though 20 to 30% of all genes encode integral membrane proteins (IMPs) [2], the structures of relatively few of these proteins have been elucidated at high resolution. Well-established expression hosts like *Escherichia coli*, yeast (*Pichia pastoris* and *Saccharomyces cerevisiae*) and higher eukaryotic cells (mammalian and insect cells) are often used for membrane protein production [3,4]. However, the production of proteins in a functional state and in sufficient yields for structural analysis is often a problem. Emerging systems like cell-free protein expression offer interesting possibilities but producing the protein in the native state is still a bottleneck [5]. Over the past decade, the Gram-positive bacterium *Lactococcus lactis* has emerged as an alternative host for membrane protein production [6-12]. *L. lactis* is amenable to genetic manipulation and well-tunable promoter systems are available [9,13,14]. The organism shows a limited proteolytic activity and, as a Gram-

positive bacterium, contains a single membrane with a high fraction of glycolipids. The easier targeting of proteins to the single (cytoplasmic) membrane, compared to Gram-negative bacteria, facilitates activator/inhibitor studies of expressed proteins [15,16]. In addition, the limited number of endogenous transporters facilitates complementation studies.

The choice of host cells for production of recombinant membrane protein depends on various factors, like gene source (codon bias, tRNA levels), protein complexity and the requirements for a particular folding environment, post-translational modifications, and others. Production of proteins can often be improved by trial-and-error approaches to screen for the best promoter, inducer levels and growth media [17,18] or by screening a wide variety of homologues [19,20]. Alternatively, production levels can be increased by selecting strains with improved protein production potential [21,22] or by screening for stable variants of a given protein [23]. To understand and ultimately alleviate the bottlenecks in membrane protein production, one needs to comprehend the response of the host cells, as has been done for yeast [24] and *E. coli* [25]. This knowledge leads to a better understanding of the cellular bottlenecks affecting the production of membrane proteins and hints towards strategies to engineer strains [26,27].

Here, we used transcriptomics and proteomics approaches to determine the response of *L. lactis* to the production of membrane proteins. We compared target membrane proteins that were either produced well or poorly, yielding different levels of stresses to the cells. We also included a secreted protein as a reference that equally requires the translocation machinery, but with potentially little or no effect on the integrity of the cytoplasmic membrane.

Results

Recombinant protein production and cell growth

The genes of the following proteins were cloned in the lactococcal nisin-inducible gene expression plasmid pNZ8048 [28] and introduced in *Lactococcus lactis* NZ9000: the ABC transporter OpuA from *L. lactis*, the sucrose transporter StSUT1 from *Solanum tuberosum*, the human PS1Δ9, a presenilin variant missing exon 9, and the extracellular α-amylase AmyQ from *Bacillus amyloliquefaciens*. Each of the proteins was engineered to contain a C-terminal hexa-histidine tag, facilitating their detection by immunoblotting. *L. lactis* NZ9000 containing the empty vector pNZ8048 was used as a control. Cells were cultured semi-aerobically at 30ºC in GM17 medium and protein expression was induced in the mid-exponential phase of growth (OD_{600} ≈ 0.5) with 10^{-3} volume of spent medium from *L. lactis* NZ9700, corresponding to final nisin A concentration of 10 µg/L. The addition of the inducer had a slight effect on growth, even in cells carrying the empty vector. A moderate effect on cell growth was observed in cells producing OpuA and AmyQ, while growth of cells producing StSUT1 and PS1Δ9 was affected more severely (Fig. 1A).

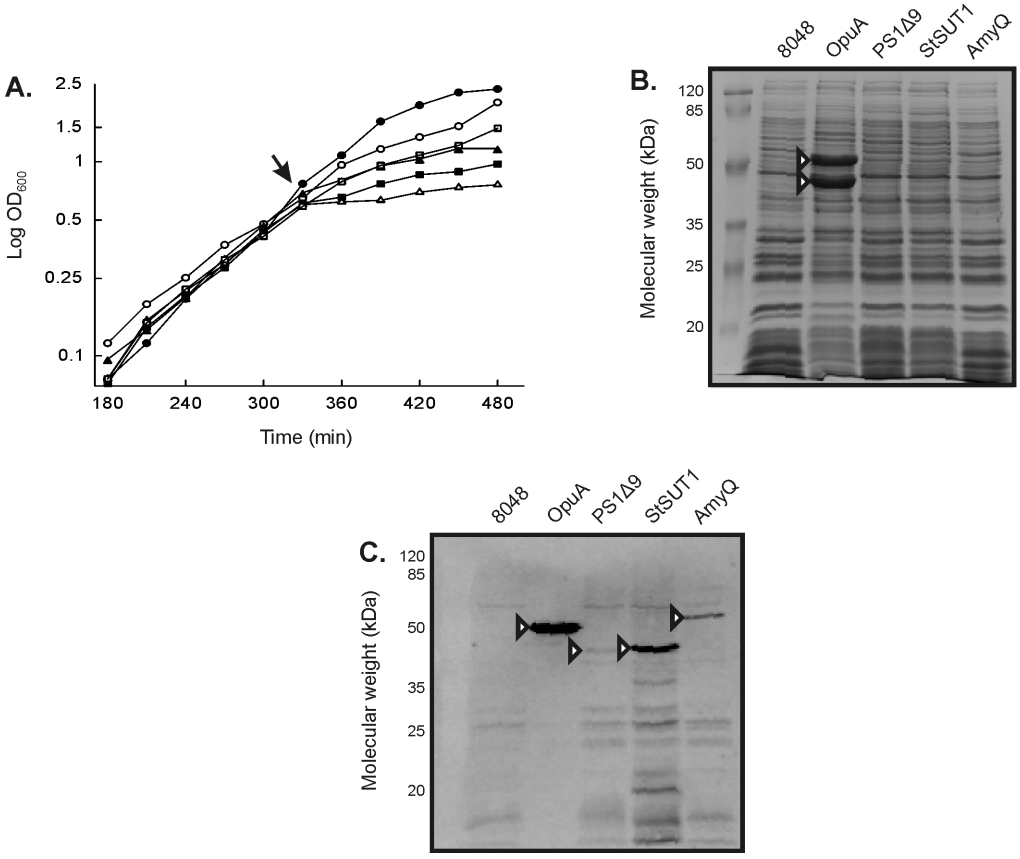


Figure 1. Characterization of membrane protein production in *L. lactis* NZ9000. Cells were grown to OD_{600} ≈ 0.5 and induced with 10 µg/L of nisin A. (A) Growth prior and after induction with nisin A (indicated by arrow) of cells producing recombinant OpuA (filled triangles), PS1Δ9 (open triangles), StSUT1 (closed squares) and AmyQ (open squares). Cells harboring the empty vector pNZ8048, induced (open circles) and uninduced (closed circles) were used as control. (B & C) Protein production in *L. lactis* NZ9000 cells. The nisin A-induced cells were harvested 2 h after induction. Protein production was analyzed on SDS-PAGE gels stained with Coomasie brilliant blue (B) and on an immunoblot using an anti-His tag antibody (C). 25 µg of total cell lystate was loaded onto the gel, except for the OpuA-expressing cells (5 µg of cell lysate to prevent saturation of the immunoblot signal). Relevant protein bands are indicated by arrow heads.

To determine protein production levels, the membrane fractions extracted from these cells, harvested 2 h after induction, were analyzed on SDS-PAGE gels. Analysis of the Coomassie-stained gel (Fig. 1B) and the immunoblot (Fig. 1C) showed very high levels of OpuA. Similar to previous studies [29], the levels of the individual subunits of OpuA (OpuAA and OpuABC) were estimated to be >10% of total membrane protein. In contrast, the production levels of the eukaryotic membrane proteins PS1Δ9 and StSUT1 were at least an order of magnitude lower. Both these proteins were only detectable on the immunoblot (Fig. 1C). AmyQ, which contains its native signal sequence for secretion into the medium, was also present in the membrane fraction, as evidenced by the immunoblot (Fig. 1C). Cell fractionation studies indicated that AmyQ was present in the cytoplasm, the membrane and in the culture medium (data not shown). This suggests that processing of the protein was incomplete, resulting in non-secreted and membrane-anchored species.

To rule out a possible bottleneck at the level of transcription of the PS1Δ9 and StSUT1 genes, RT-qPCR was performed to determine the mRNA levels of *PS1Δ9*, *StSut1* and *opuA*. We observed that 64 min after induction the fold increase of the two mRNA species were comparable to that of the *opuA* transcript (Fig. 2D), thus indicating that transcription is not a limiting factor for the production of the two eukaryotic proteins.

Strains and constructs for "Omics" analysis

To probe the basis for the difference in the production levels of the four proteins under study, we determined the physiological response of *L. lactis* to the protein synthesis burden by performing a combined proteomic and transcriptomic analysis. Contrary to PS1Δ9, StSUT1 and AmyQ, that have no known activity in *L. lactis* and/or of which the substrates are not present in our experimental setup, increased production of OpuA results in the accumulation of glycine betaine and, therefore, in an increase in the internal osmotic pressure of cells. In order to be able to only examine the effect of overproduction of the OpuA protein, we expressed a point-mutated version of OpuA, OpuA(H223A), in which the histidine residue at position 223, essential for the ATPase activity [30], was replaced by alanine.

OpuA is easily produced by *L. lactis* NZ9000 and is therefore expected to request a considerable amount of resources, for transcription as well as for translation and translocation. To eliminate the putatively significant transcription burden from our studies, we also made pNZ*opuA*mRNA in which the AUG codons at positions 1 (start of *opuAA*), 1224 (start of *opuABC*), and 1344 (potential internal start site), were substituted by UAA stop codons. This plasmid was used in all subsequent experiments as a control for the production of OpuA(H223A). Fig. 2A shows that OpuA and OpuA(H223A) were synthesized at comparable levels in *L. lactis* and that no OpuA was formed in *L. lactis* NZ9000/pNZ*opuA*mRNA. Whole-cell ^{14}C-glycine betaine transport assays, using *L. lactis* Opu401 [a NZ9000 derivative lacking the chromosomal *opuA* gene [31]] carrying the proper expression plasmids, confirmed that wildtype OpuA was functionally expressed and that OpuA(H223A) was inactive (Fig. 2B). Induced expression of wildtype OpuA and OpuA(H223A) had similar negative effects on the growth of *L. lactis* (Fig. 2C). Growth of the mRNA control strain (*L. lactis*(NZ9000/pNZ*opuA*mRNA) was only slightly inhibited upon induction with nisin A and similar to that of the of empty vector control strain NZ9000/pNZ8048. These data suggest that the nisin A-induced increase of transcription *per se* does not have a major impact on the physiology of *L. lactis*. Growth of the producer organism is inhibited only when either the functional or the inactive form of OpuA is produced to high levels.

Thus, to characterize the response of *L. lactis* to membrane protein production stress, proteomics and transcriptomics studies were performed with the *L. lactis* NZ9000 strains carrying pNZOpuA(H223A), pNZPS1Δ9, pNZStSUT1 or pNZAmyQ. *L. lactis* NZ9000/pNZ*opuA*mRNA was used as a dedicated control for *L. lactis*

NZ9000/pNZOpuA(H223A), whereas *L. lactis* NZ9000/pNZ8048 was used for the same purpose in the three other situations.

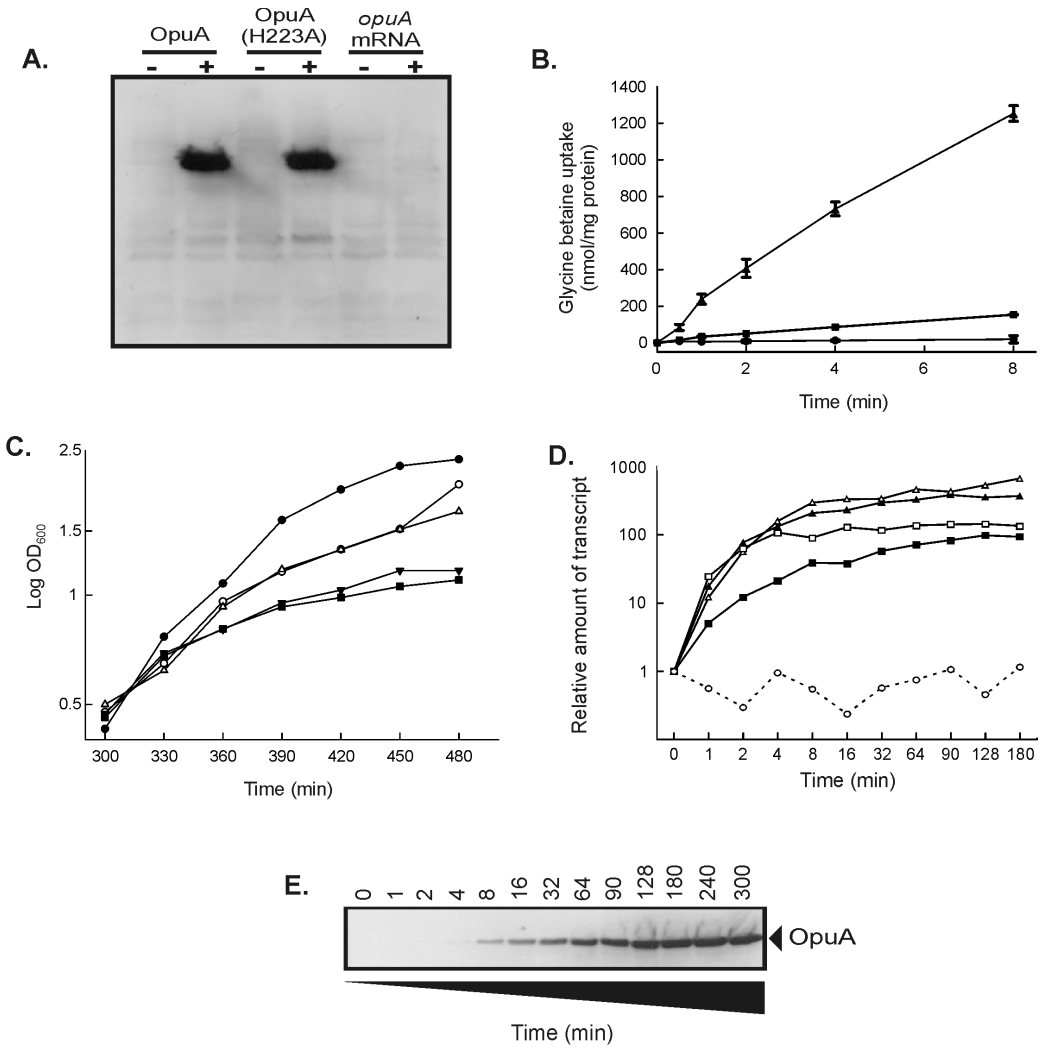


Figure 2. Transcript, protein and activity levels. (A) The recombinant production of proteins in *L. lactis* NZ9000 was analyzed on immunoblot, using an anti-His tag antibody ("–" indicates uninduced cells and "+" cells induced with 10 µg/L of nisin A at OD_{600} ≈ 0.5). Cells were harvested after 2 h of nisin A induction. (B) Transport activity of parental OpuA (triangles) and OpuA(H223A) (squares) was analyzed in *L. lactis* Opu401 cells. The nisin concentration for induction was 100 ng/L and the uninduced cells, carrying the pNZ8048 (circles), were used as a control. (C) Growth of *L. lactis* NZ9000, expressing parental OpuA (filled triangles), OpuA(H223A) (filled squares) and *opuA*mRNA (empty triangles). Cells harboring pNZ8048 (empty vector), induced (open circles) and uninduced (closed circles) were used as control. (D) Time-resolved transcript profiling of *opuAA* (filled triangles), *opuABC* (empty triangles), *cesSR* (filled squares), *dnaK* (empty squares) and *bcaP* (empty circles) after induction, as analyzed by RT-qPCR. (E) Time-resolved production of parental OpuA protein by *L. lactis* NZ9000 cells, as analyzed on immunoblot with an anti-histidine tag antibody.

Transcription and Translation Kinetics/Dynamics

To optimize the sampling for the proteomics and transcriptomics studies, we determined the time-dependent profiles of transcription of *opuA* and a number of genes

involved in the stress response (Fig. 2D) as well as the synthesis of the OpuA protein (Fig. 2E). Fig. 2D shows that transcripts for the individual subunits of OpuA (OpuAA and OpuABC) can be readily detected after induction, rising quickly and reaching a plateau after about 15 min of induction. The transcripts levels of *cesR* and *dnaK*, indicators of a specific and a general stress response in *L. lactis*, respectively (see accompanying paper), were determined in parallel. The kinetics of formation of these transcripts was similar to that of *opuA*.

Expression of branched chain amino acid permease (*bcaP*) remained constant throughout the growth curve [32] and was used as a control. Although increased transcript levels could be detected after one min of induction, it took longer to observe the overproduced proteins by Western blotting (Fig 2E). OpuABC was detected after 4 to 8 min of induction and increased gradually over a period of 120 min. Therefore, sampling for transcriptomics was dense (0, 2, 8, 16, 32 and 64 min after induction), but less frequent for proteomics, (0 min, 16 min (onset of detectable OpuA) and 64 min (high accumulation of OpuA)). Time points later than 64 min were not examined as many indirect physiological phenomena were expected to obscure the (initial) response to the overproduction of the membrane proteins.

Time-resolved stress response and experimental approach

L. lactis was grown in triplicate in pH-regulated bioreactors. Cell samples were collected in methanol at -40^{o}C to immediately quench further synthesis or breakdown of mRNA. For the proteomic studies, translation in harvested cells was stopped immediately by adding 100 µg/mL of chloramphenicol. Transcriptome data was acquired by hybridizing the dual dye-labelled cDNA, obtained by reverse transcription of purified mRNA, onto SuperAmine glass slides spotted with duplicates of around 2500 ORF amplicons from *L. lactis* subsp. *cremoris* MG1363 [33]. Principal component analysis (PCA) of the transcriptome data showed that the cells reacted immediately on the induction of the production of OpuA(H223A), using NZ9000/pNZ*opuA*mRNA as the reference, and significant differential expression of genes was observed already after 8 min. After the initial response the PCA analysis showed a gradual change in the responses up to the 32 and 64 min time point samples (Fig. 4A). After preliminary analysis of the transcriptome data, the 64 min time point seemed to be most relevant and, for pragmatic reasons, was therefore selected for the transcriptomics analysis in the experiments comparing *L. lactis* strains NZ9000/pNZPS1Δ9, NZ9000/pNZStSUT1 and NZ9000/pNZAmyQ.

The soluble proteome of the control cells (NZ9000/pNZ8048 or NZ9000/ pNZ*opuA*mRNA) and cells overproducing membrane protein was analyzed by two-dimensional differential in gel electrophoresis (2D-DIGE), in combination with nanoLC-MS/MS (see Materials and Methods). Upon OpuA(H223A) overproduction, the PCA analysis on the spot volumes, for a subset of protein spots that were matched across all the 2D gel images (Decyder 6.5 GE Healthcare, Uppsala, Sweden), revealed no significant differences for the time points 0 min and 16 min. Differences between the induced *L.*

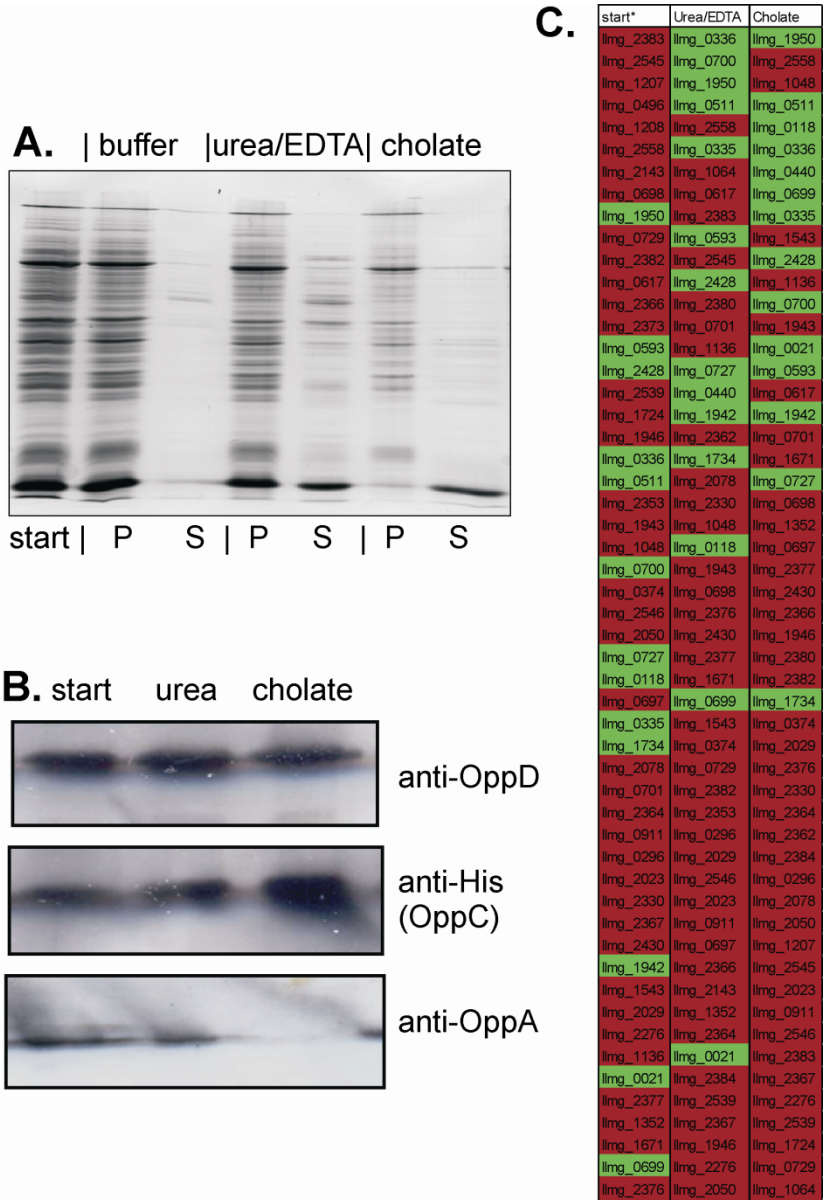
lactis NZ9000/pNZ*opuA*mRNA and NZ9000/pNZOpuA(H223A) cultures became significant only after 64 min. The soluble proteomes were then distinct from each other and from the earlier time points (Fig. 4B) and were further analyzed.

Membrane proteomes extracted from control cells (NZ9000/pNZ8048 or NZ9000/ pNZ*opuA*mRNA) and cell overproducing membrane protein collected after 64 min of induction were analyzed by 2D-liquid chromatography, separating iTRAQ-labelled peptides by strong cation exchange (SCX) and reverse phase nano-liquid chromatography (RP-nLC). Proteins were identified by MALDI-MS/MSMS of the peptides. To improve the identification and quantification of low abundant membrane proteins, the membrane fractions were extracted with urea/K-EDTA and subsequently with cholate to remove the majority of aspecifically bound cytosolic proteins (for more detail, see paragraph below). Nevertheless, highly abundant soluble cytosolic proteins (glycolytic enzymes, ribosomal proteins) could still be detected in the membrane proteome fraction and these were included in the proteome dataset. The significant differentially expressed proteins and all identified proteins (crude data) are listed in Supplementary Table SI.

Extraction of the membrane vesicles

Prior to analysis of the membrane proteome via iTRAQ labeled peptides, membrane vesicles were extracted in two steps using urea/EDTA and cholate to remove loosely attached soluble proteins from this fraction and to improve detection of the membrane proteins. Urea and cholate extractions have been described previously for the lactose carrier [34] and EDTA was used for the stripping of membrane containing H^+-ATPases [35]. The SDS-PAGE gel (Fig. 3A) shows the differences for the membranes after washing with buffer, with urea/K-EDTA and with cholate after urea/K-EDTA washing. Washing with buffer was not an efficient extraction method, while washing with urea/K-EDTA shows a substantial amount of proteins in the supernatant, also with a different pattern compared to the membrane fraction. The pattern of the membrane fraction is altered significantly after subsequent washing with cholate compared both to the startmaterial and the urea/EDTA extracted sample. Small scale MALDI-MS/MSMS analysis of these fractions show that upon first identification of the start fraction 23% of the identified proteins are ribosomal proteins, which remains equal at around 25% for the urea/K-EDTA fraction (but more proteins are identified in total) and goes down to 4% after the subsequent washing with cholate, showing the strong effect of these extractions on soluble proteins. The specificity to remove the soluble proteins only was shown using membrane vesicles containing the over produced membrane protein Opp, which could be detected using protein-specific antibodies (Fig. 3B). These results show that membrane domains of the proteins are retained during the extraction steps. Relative quantification of the membrane extraction effects were analysed using iTRAQ labeling. Start material was compared to start material dissolved in a different buffer (start*), the urea/EDTA extraction and to the cholate extraction subsequently after the urea/EDTA

extraction. The identified proteins were sorted to their iTRAQ ratio's, meaning that proteins in top of the list are enriched relatively compared to the startmaterial as shown in Fig. 3C.



start*	Urea/EDTA	Cholate
llmg_2383	llmg_0336	llmg_1950
llmg_2545	llmg_0700	llmg_2558
llmg_1207	llmg_1950	llmg_1048
llmg_0496	llmg_0511	llmg_0511
llmg_1208	llmg_2558	llmg_0118
llmg_2558	llmg_0335	llmg_0336
llmg_2143	llmg_1064	llmg_0440
llmg_0698	llmg_0617	llmg_0699
llmg_1950	llmg_2383	llmg_0335
llmg_0729	llmg_0593	llmg_1543
llmg_2382	llmg_2545	llmg_2428
llmg_0617	llmg_2428	llmg_1136
llmg_2366	llmg_2380	llmg_0700
llmg_2373	llmg_0701	llmg_1943
llmg_0593	llmg_1136	llmg_0021
llmg_2428	llmg_0727	llmg_0593
llmg_2539	llmg_0440	llmg_0617
llmg_1724	llmg_1942	llmg_1942
llmg_1946	llmg_2362	llmg_0701
llmg_0336	llmg_1734	llmg_1671
llmg_0511	llmg_2078	llmg_0727
llmg_2353	llmg_2330	llmg_0698
llmg_1943	llmg_1048	llmg_1352
llmg_1048	llmg_0118	llmg_0697
llmg_0700	llmg_1943	llmg_2377
llmg_0374	llmg_0698	llmg_2430
llmg_2546	llmg_2376	llmg_2366
llmg_2050	llmg_2430	llmg_1946
llmg_0727	llmg_2377	llmg_2380
llmg_0118	llmg_1671	llmg_2382
llmg_0697	llmg_0699	llmg_1734
llmg_0335	llmg_1543	llmg_0374
llmg_1734	llmg_0374	llmg_2029
llmg_2078	llmg_0729	llmg_2376
llmg_0701	llmg_2382	llmg_2330
llmg_2364	llmg_2353	llmg_2364
llmg_0911	llmg_0296	llmg_2362
llmg_0296	llmg_2029	llmg_2384
llmg_2023	llmg_2546	llmg_0296
llmg_2330	llmg_2023	llmg_2078
llmg_2367	llmg_0911	llmg_2050
llmg_2430	llmg_0697	llmg_1207
llmg_1942	llmg_2366	llmg_2545
llmg_1543	llmg_2143	llmg_2023
llmg_2029	llmg_1352	llmg_0911
llmg_2276	llmg_2364	llmg_2546
llmg_1136	llmg_0021	llmg_2383
llmg_0021	llmg_2384	llmg_2367
llmg_2377	llmg_2539	llmg_2276
llmg_1352	llmg_2367	llmg_2539
llmg_1671	llmg_1946	llmg_1724
llmg_0699	llmg_2276	llmg_0729
llmg_2376	llmg_2050	llmg_1064

Figure 3. Small scale analyses of the extraction of membrane vesicles. A) SDS-PAGE analysis of the extraction of membrane vesicles. Samples were compared after extraction with buffer, after urea/K-EDTA extraction and urea/K-EDTA plus subsequent cholate extraction. The pellet fraction (p) contains the membrane vesicles and the supernatant (s) contains the proteins that were extracted from the membrane vesicles. B) Immunoblot analysis of the extraction with membrane vesicles containing the overproduced membrane protein Opp. Detection was done against the ATPase (OppD), the transmembrane domain (OppC) and the lipid anchored substrate binding domain (OppA). C) Quantified proteins using iTRAQ labeling were sorted by iTRAQ ratio (ratio of extracted sample over start material, sorted from high (relative enrichment) to low (relative depletion)). Green colored protein names contain at least one transmembrane helix (based on TMHMM predictions) and red colored proteins do not.

The use of different buffers (start*) did not improve the identification of the membrane proteins (green colored proteins), which can be seen by the random distribution of soluble proteins and membrane proteins in Fig. 3C.

Both extraction methods show a significant clustering of most of the membrane proteins in the top of the list, indicating enrichment of the membrane proteins compared to the startmaterial. The results are even better if you keep in mind, that proteins listed as soluble proteins, can still be attached to the membrane via covalent binding to membrane proteins or via lipid anchoring and are part of membrane protein complex (for example the substrate binding domain of many transporters in the membrane), which is not taken into account in this list.

The results show a clear improvement in the identification of the membrane proteins on small scale experiments and in the full proteome study presented in this paper 37% of the identified proteins contained at least one transmembrane helix. Similar full proteome analyses done for the membrane proteome of the human ABC transporter CFTR, using the membrane vesicles directly without applying any extra membrane extraction steps contained on average 22% of proteins with at least one transmembrane helix [36]. Since all other conditions (iTRAQ labeling, 2D chromatography and MALDI-MS/MSMS analysis) were similar, these results show that the optimized detection of membrane proteins observed in the small scale experiments also holds true for full scale proteome analyses.

The physiological response of L. lactis to membrane protein overproduction

The physiological response of *L. lactis* to recombinant protein overproduction was determined by analyzing the changes in the transcriptome and the proteome and by comparing these datasets (Table SI). A global quantification of the statistically significant observations is presented in Table 1. More changes were observed at the level of the transcriptome than at the proteomic level. At the transcriptome level, most changes were seen upon overproduction of OpuA(H223A), whereas overproduction of the three other proteins yielded about the same number of differentially expressed genes (Table 1). At the proteome level, more changes were observed upon overproduction of OpuA(H223A) and StSUT1 than for PS1Δ9 and AmyQ (Table 1). The correlation between the transcriptome and proteome datasets are represented as Venn diagrams in Fig. 4C, showing that for about ~50% of the identified differentially expressed proteins the corresponding genes in the transcriptome were identified to be differentially regulated. The only exception was the proteome dataset obtained after StSUT1 overproduction, where only 30% of the proteome dataset showed correlation to the transcriptome data. The Venn diagrams presented in Fig. 4D & 4E show the overlap between the different transcriptome (Fig. 4D) and proteome (Fig. 4E) datasets. With respect to the effects on gene expression, the strains overproducing membrane proteins showed more overlap to each other than to the AmyQ overproducer, with the latter being most close to the strain producing OpuA(H223A). The clustering of strains overproducing the three

membrane proteins was also observed in the proteome data. From the perspective of the proteome, overproduction of AmyQ seems to relate most to StSUT1 overproduction.

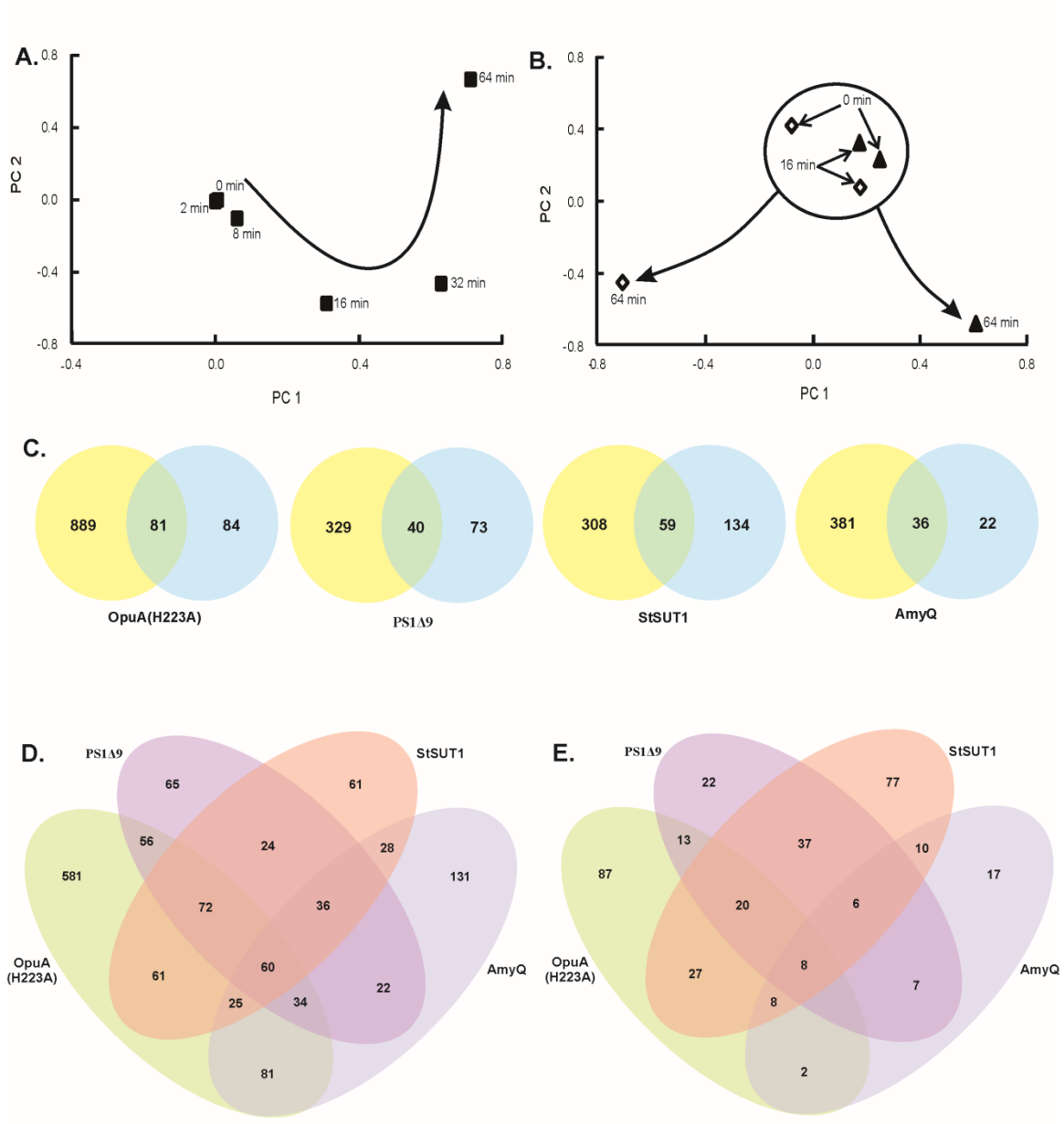


Figure 4. Principal component analysis (PCA) of transcriptome (A) and proteome data obtained from 2D gel spot maps (B) of cells producing OpuA (H223A) (full triangles) and compared with the *opuA*mRNA control (open diamonds). The 2D gels were analyzed and spot volumes were extracted using Decyder 6.5 (GE Healthcare, Uppsala, Sweden). Transcriptome PCA (full squares) was obtained from the ratio data of the two populations of cells. (C) Overlap in the transcriptome (yellow) and proteome data (blue). The number of overlapping genes/proteins are indicated in the overlap of both circles. Overlap of significant differences in the transcriptome (D) and proteome (E) datasets obtained for OpuA(H223A) (Green), PS1Δ9 (purple), StSUT1 (pink) and AmyQ (Blue).

Table 1. Number of up- and down-regulated genes/proteins in *L. lactis* strains.

Comparison	**Micro-arrays**		**2D DIGE**		**iTRAQ**	
	Up	**Down**	**Up**	**Down**	**Up**	**Down**
pNZ*opuA*(mRNA) *versus* pNZOpuA(H223A)	460	510	12	9	9	142
pNZ8048 *versus* pNZPS1Δ9	182	187	9	21	35	52
pNZ8048 *versus* pNZStSUT1	193	174	7	13	113	63
pNZ8048 *versus* pNZAmyQ	144	273	5	5	27	25

L. lactis NZ9000, carrying the indicated plasmids, were grown as described under experimental procedures and numbers of differentially expressed genes and proteins were determined.

Taking the data as a whole, it is clear that (over)production of the membrane proteins OpuA(H223A), PS1Δ9, StSUT1 elicited a bigger response in *L. lactis* than did AmyQ production (Table 1). The stress response was most pronounced for cells overproducing OpuA(H223A), which can be explained by the fact that this protein is produced to much higher levels than PS1Δ9 or StSUT1. We also note that growth of *L. lactis* was inhibited upon induction of the synthesis of OpuA(H223A), PS1Δ9, StSUT1 and AmyQ, and that a large part of the response may reflect adaptation to the lower growth rate. Some remarkable differences were observed in the physiological response of *L. lactis* upon production of the membrane proteins OpuA(H223A), PS1Δ9 and StSUT1 and the secreted protein AmyQ. These are discussed in the sections below:

1) General stress response

The proteome and transcriptome data show that recombinant membrane protein production in *L. lactis* evoked a so-called general stress response, including the up-regulation of *hrcA-grpE-dnaK*, *dnaJ*, *groES-groEL*, *clpP*, *clpB*, *clpE* and *clpC* (Fig. 5A). This response is a clear indication that the pool of mis-folded proteins is increased. Possibly, some of the produced protein molecules might not have correctly assembled in the membrane, *e.g.* due to an overload of the membrane targeting and/or the translocation machinery. The identification of OpuAA, the nucleotide-binding subunit of the ABC transporter OpuA, in the soluble proteome confirms that a significant fraction of the protein does not become membrane-associated upon overproduction (Table SI). Contrary to what was observed with the membrane proteins, in case of AmyQ overproduction none of the genes/proteins (except DnaK at proteome level) concerned with the general stress were differentially regulated. Interestingly both at the proteome and transcriptome level HtrA, a protein involved in quality control of secretory proteins was identified to be upregulated, which could be due to the accumulation of misfolded protein in the cell envelope (Fig. 5A).

2) Carbon and energy metabolism

The levels of transcripts encoding glycolytic enzymes and pyruvate-dissipating enzymes were decreased in the OpuA(H223A), PS1Δ9, StSUT1 and AmyQ producing cells. Similar observations were made at the proteome level (Fig. 5B). Like for nucleotide metabolism (discussed below), the trends correlate with the decrease in growth rate and may reflect

the lower need for metabolic energy. The transcript levels for the subunits of F1F0-ATPase (*atpA*, *B*, *D*, *E*, *F*, *G* and *H*) were higher in cells overproducing OpuA(H223A), even though AtpD and A were down-regulated at the protein level. In strains overproducing PS1Δ9 or StSUT1, AtpD, A and H were down-regulated at the protein level (Fig. 5B).

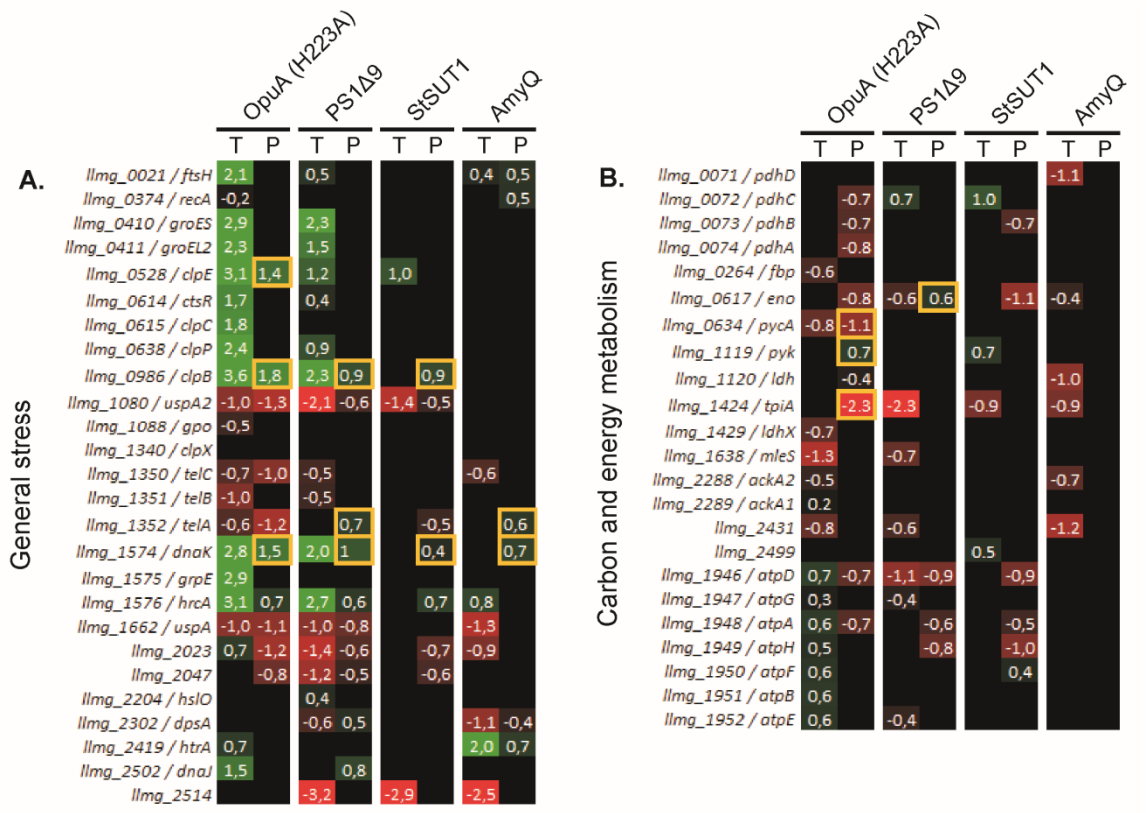


Figure 5. Quantification of mRNA and protein differences from the general stress (panel A) and carbon and energy metabolism (panel B) in cells producing recombinant proteins. The depicted values are $\log_2$-transformed linear ratios of the average data obtained from biological triplicates. "T" stands for transcriptome data and "P" for membrane proteome (iTRAQ) data, except when a yellow bordered square is depicted, in which case data was obtained by 2D DIGE. Only statistically significant *p*-values are indicated; the color intensity is a measure of the depicted ratios (green for up-regulation and red for down-regulation).

3) Nucleotide metabolism

A sharp down-regulation of genes involved in the synthesis of nucleotides (purines and pyrimidines) via the *de novo* and salvage pathways was observed in all samples. Furthermore, genes encoding proteins involved in nucleotide degradation (*cdd* and *udp*) and synthesis of deoxy-ribonucleotides (*nrdI, E, F and llmg_0281*) were down-regulated, but most of these changes were not observed at the protein level (Fig. 6A). The responses of nucleotide metabolism to a variety of environmental stresses has been observed previously and is most likely not specific for membrane protein production. Most likely, it relates to the reduced growth rate of *L. lactis*, which has previously been coupled to a down-regulation of *pyr* and *pur* genes [37-39].

4) Amino acid biosynthesis

The levels of transcripts encoding for enzymes involved in amino acid biosynthesis, except for methionine (MetC) and cysteine (CysD/K), were increased, but these effect were not seen at the proteome level. Up-regulation of these transcripts was highest for the OpuA(H223A)-producing cells (Fig. 6B). The probable reduction of the cytoplasmic concentration of free amino acids, due to the significant overproduction of OpuA, might result in the upregulation of these genes. In addition, the transcripts for the cytosolic peptidases (*pepO*, *N*, *C* and *F*) were also increased upon recombinant protein expression. The *pepN* gene is disrupted in *L. lactis* NZ9000 by the integration of *nisRK,* but its regulation remains preserved. Induction of *pepN* can still be measured as the array probe does not overlap with the integrated fragment containing *nisRK.* The expression of all these genes is controlled by a global nitrogen metabolism regulator, CodY, whose repressing effect is relieved when the intracellular concentration of branched-chain amino acids becomes limiting [40]. Thus, the differential regulation of the peptidase genes is consistent with a limitation in (branched-chain) amino acids in OpuA(H223A)-producing cells (see also [41]). Notably, the genes for the di-peptide and oligo-peptide transporters (*dpp*, *dtpT* and *opp*) genes, which are also regulated by CodY, were down-regulated (Fig. 6B). This suggests that for membrane proteins the possible CodY-mediated upregulation may be overruled by a specific cell envelope stress response (see below).

5) Protein synthesis

A severe effect on the cell's translational machinery was observed upon production of all 4 recombinant proteins, as was the case for ribosomal proteins, both at the level of the transcriptome and the proteome (Fig. 7). Translation and the tRNA synthetase proteins/genes show more mixed effects, which were not specific for any of the samples (Fig. 7). Interestingly, almost all of the differentially-regulated ribosomal proteins were observed in the membrane proteome (despite the urea/EDTA and cholate treatment) but not in the soluble proteome, which could mean that ribosomal proteins are more tightly interacting with (proteins in) the membrane suggesting a coupled transcription and translocation process.

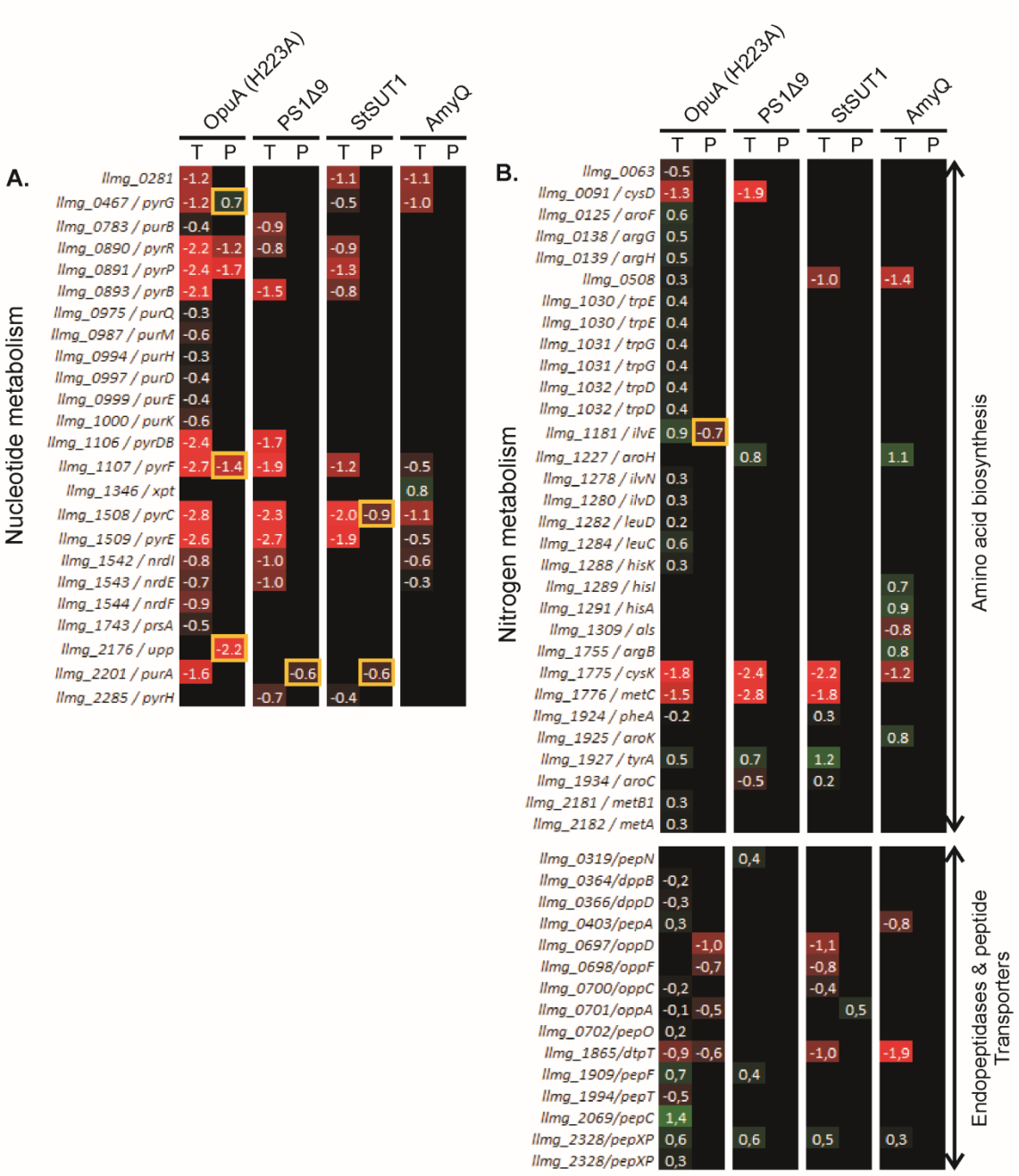


Figure 6. Quantification of mRNA and protein differences from the nucleotide metabolism (panel A) and nitrogen metabolism (panel B) in cells producing recombinant proteins. The depicted values are $\log_2$-transformed linear ratios of the average data obtained from biological triplicates. "T" stands for transcriptome data and "P" for membrane proteome (iTRAQ) data, except when a yellow bordered square is depicted, in which case data was obtained by 2D DIGE. Only statistically significant *p*-values are indicated; the color intensity is a measure of the depicted ratios (green for up-regulation and red for down-regulation).

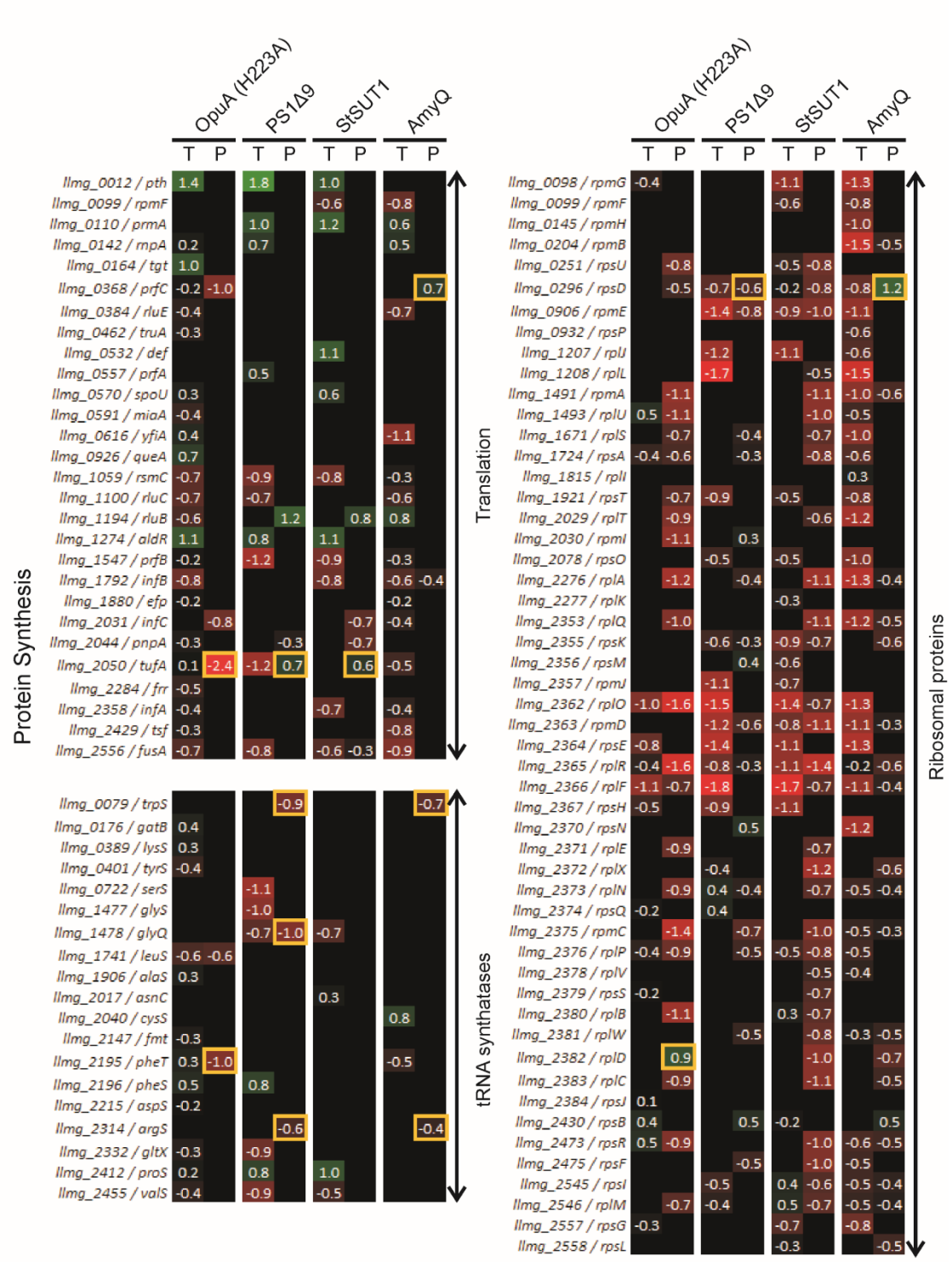


Figure 7. Quantification of mRNA and protein differences from the protein synthesis in cells producing recombinant proteins. The depicted values are $\log_2$-transformed linear ratios of the average data obtained from biological triplicates. "T" stands for transcriptome data and "P" for membrane proteome (iTRAQ) data, except when a yellow bordered square is depicted, in which case data was obtained by 2D DIGE. Only statistically significant *p*-values are indicated; the color intensity is a measure of the depicted ratios (green for up-regulation and red for down-regulation).

6) Cell envelope biosynthesis

The biosynthesis of the peptidoglycan layer and of lipoteichoic acids showed similar responses (upregulation of many of the proteins and transcript levels), independent of the overproduced protein. A few protein-specific trends were observed. For instance, four of the penicillin-binding proteins (Pbp1A/2A/2B/X) were up-regulated specifically in the proteome of cells producing StSUT1. Together with the up-regulation of genes involved in the biosynthesis of lipoteichoic acids and the D-alanylation of these acids, these data indicate an alteration of the mechanical properties of the cell wall. With respect to lipid synthesis, the transcriptomic and proteomic changes point towards an increase in the concentration of phosphatidylglycerol and a decrease in cardiolipin upon overexpression of OpuA(H223A), PS1Δ9 and StSUT1 (Fig. 8A). An increase in cardiolipin at the expense of phosphatidylglycerol is often linked to stress-conditions and/or a reduced growth [42]. Down-regulation of the fatty acid synthesis genes (the *acc* and *fab* genes) was observed at the transcript level upon overproduction of OpuA(H223A), PS1Δ9 and StSUT1 but not of AmyQ. The two genes that are differentially expressed in the AmyQ producing cells, *i.e., fabG* and *fabZ*, are in fact up-regulated.

7) Cell envelope stress response and protein translocation

A number of genes/proteins that were upregulated in response to the overproduction of proteins, such as *ftsH, oxaA2, ppiB, pacL* and *telA*, form part of the CesSR regulon and constitute a cell envelope stress response [43,44] (Fig. 8B). Some of these proteins, notably FtsH and OxaA2, are known to play crucial roles in membrane protein biogenesis [45]. Their upregulation may increase the capacity to remove misfolded protein (FtsH, a membrane-bound cell division protease), while allowing a more efficient insertion or folding of proteins into the membrane (OxaA2). An increased capacity of protein translocation is furthermore suggested by the upregulation of Ffh and SecG (Fig. 8B), albeit that the SecY component of the translocon was not affected in its expression. The differential expression of all of these genes was seen with all 4 overproduced proteins, indicating that both AmyQ and the integral membrane proteins induce a CesSR response.

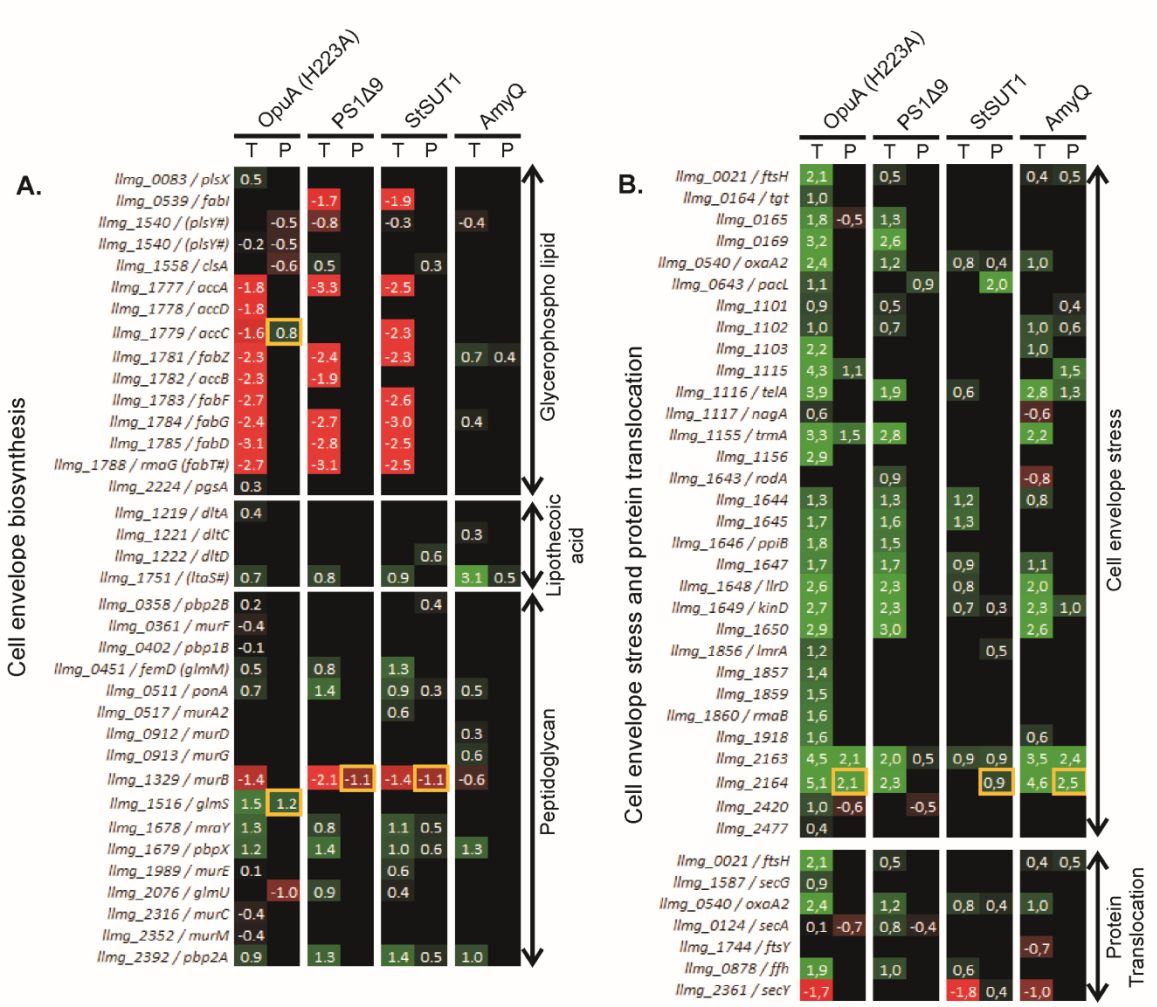


Figure 8. Quantification of mRNA and protein differences from the cell envelope biosynthesis (panel A) and cell envelope stress and protein translocation (panel B) in cells producing recombinant proteins. The depicted values are $\log_2$-transformed linear ratios of the average data obtained from biological triplicates. "T" stands for transcriptome data and "P" for membrane proteome (iTRAQ) data, except when a yellow bordered square is depicted, in which case data was obtained by 2D DIGE. Only statistically significant *p*-values are indicated; the color intensity is a measure of the depicted ratios (green for up-regulation and red for down-regulation).

8) Membrane transport

The inorganic ion transporters showed mostly mixed effects upon overexpression of the different membrane proteins and only a subgroup showed consistent effects within a category (Fig. 9). The vitamin transporters (CbiO3/O2/Q2, PanT, NiaX, QueT) showed a general down-regulation upon overexpression of the membrane proteins. The copper ATPases (CopAB) and the multidrug transporters (all ABC type transporters and about half of the major facilitator superfamily) were up-regulated in the strains overproducing OpuA(H223A), PS1Δ9 and StSUT1 but not AmyQ, which might reflect a growth-related effect. Three transporters showed a specific effect in the proteome analysis: the iron (FeoB and FhuCBD), the manganese (MtsABC) and the magnesium (MgtA-type) transporters like PacL were up-regulated upon PS1Δ9 or StSUT overproduction, down-

regulated in the strain producing OpuA(H223A) and not differently expressed in the AmyQ overproducer.

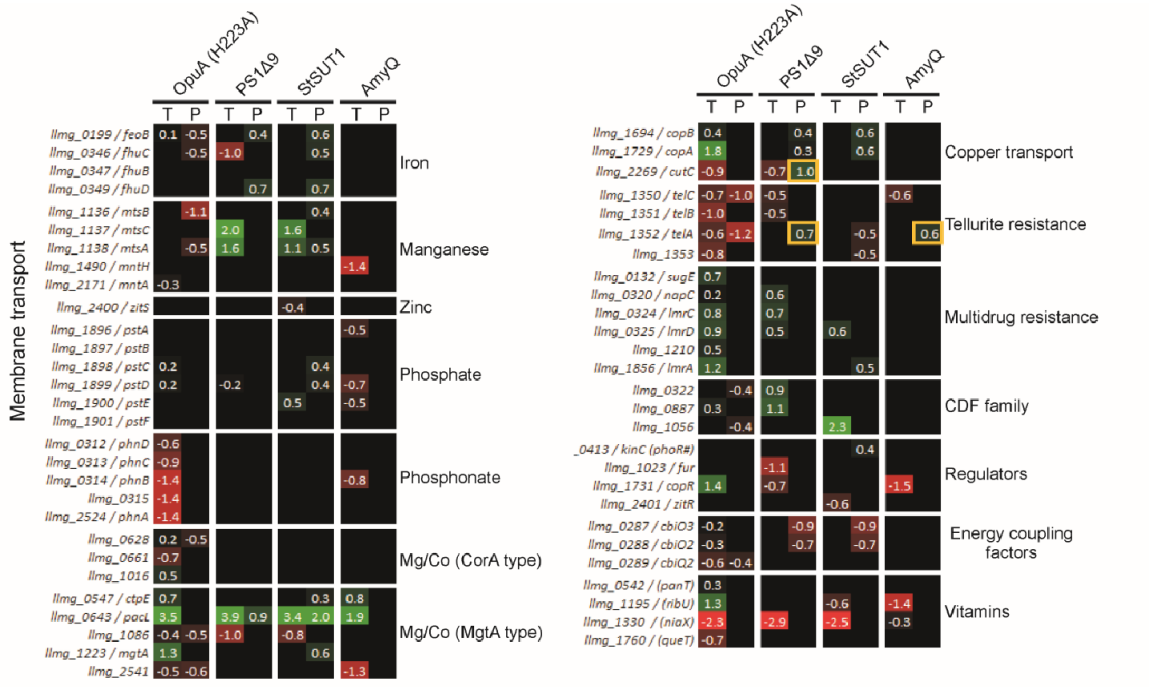


Figure 9. Quantification of mRNA and protein differences from the cell envelope biosynthesis (panel A) and cell envelope stress and protein translocation (panel B) in cells producing recombinant proteins. The depicted values are $\log_2$-transformed linear ratios of the average data obtained from biological triplicates. "T" stands for transcriptome data and "P" for membrane proteome (iTRAQ) data, except when a yellow bordered square is depicted, in which case data was obtained by 2D DIGE. Only statistically significant *p*-values are indicated; the color intensity is a measure of the depicted ratios (green for up-regulation and red for down-regulation).

Discussion

The major hurdle in the structural analysis of membrane proteins is their overproduction in a functional state. Producing membrane proteins requires coordination of several processes, such as transcription, translation, targeting, membrane insertion, folding, and, in many cases, post-translational modifications. A thorough analysis of the host cell response to and mechanistic information about membrane protein biogenesis will aid in the design of strategies to optimize the recombinant production of these proteins. Here, we report on the physiological response of *L. lactis* NZ9000 to apparent stress(es) evoked by the synthesis of a number of integral membrane proteins and of the secretory protein AmyQ. We show that, although OpuA is expressed to much higher levels than PS1Δ9 and StSUT1, the response of the cells to the stress of producing these proteins is very similar, which was not anticipated. As the amount of PS1Δ9 and StSUT1 synthesized is very low, it is unlikely that the growth inhibition is due to diversion of nutrients towards the synthesis of recombinant protein. What, then, causes the inhibition of growth? In *E. coli* and *B. subtilis* down-regulation of genes involved in transcription,

translation and tRNA synthetases, similar to what is observed here in *L. lactis*, is invoked by the synthesis of ppGpp via the ribosome-associated RelA protein in the absence of charged tRNA's. The accumulation of ppGpp acts as an alarmone, repressing the transcription of various genes essential for cell growth. This phenomenon is called the stringent stress response [46,47]. The stringent stress response in *L. lactis* has been identified in for instance acid stressed [48,49] and might play a role here.

The production of recombinant membrane protein was accompanied by a general stress response, as the heat shock proteins DnaK, GroeEL, DnaJ and GrpE were highly upregulated. The expression of these genes is controlled by HrcA, a regulatory protein that binds to the CIRCE sequence present in the upstream region of these genes [50,51]. In addition, the protease ClpE and the chaperones ClpB, E and X, whose expression is regulated by CtsR [52], were also upregulated. The response was observed for a broad range of stresses such as acid, heat and osmotic challenges [53,54]. The response is triggered by the accumulation of misfolded protein, suggesting that not all OpuA(H223A), PS1Δ9 and StSUT1 is correctly folded. Alternatively, it might be that the stress results from an increased population of generic unfolded proteins, due to depletion of the folding machinery by the overproduced recombinant proteins. Interestingly none of the proteins concerned with general stress were differentially regulated in case of AmyQ overexpression, even though only part of the produced AmyQ was secreted and, thus, presumably correctly processed.

The most important finding of our work is the CesSR-mediated response, resulting in the differential regulation of a wide variety of genes, many of which have a (putative) central role in maintaining cell envelope integrity and its normal functions, such as LmrA (a multidrug resistance ABC transporter), RmaB (a transcriptional regulator of the MarR family), SpxB, OxaA2 (membrane insertase/foldase) or FtsH (AAA-type ATP-dependent membrane-bound metalloprotease) (Fig. 8B). CesSR is a two-component regulatory system (TCS) that orchestrates a cell envelope stress response [55]. Although it is not known what is actually sensed by CesS, the histidine-kinase and sensor component of the system, the system responds to the presence of bacteriocins [56] and lysozyme [57]. We show here that CesSR senses an envelope stress evoked by the overproduction of a variety of membrane proteins, and of a secretory protein. This occurs presumably via the association of misfolded hydrophobic protein with the cytoplasmic membrane. In fact, by using GFP as a quality control indicator of correctly folded protein, we have observed that misfolded membrane proteins in *L. lactis* do not end up in electron-dense inclusion bodies (as frequently observed in *E. coli*) but, rather, are associated with the membrane lipid fraction of the cell [58]. An important role for FtsH seems obvious under those circumstances. Most genes from the CesSR regulon code putative membrane proteins or proteins acting on the cytoplasmic membrane, clearly indicating the specificity of this response. The influence of CesSR and members of this regulon on membrane protein production is described in the accompanying paper (Pinto et al. 2010).

Conclusion

By using a combined proteomic and transcriptomic approach we were able to identify the physiological response of *L. lactis* to membrane protein overproduction (Fig. 10).

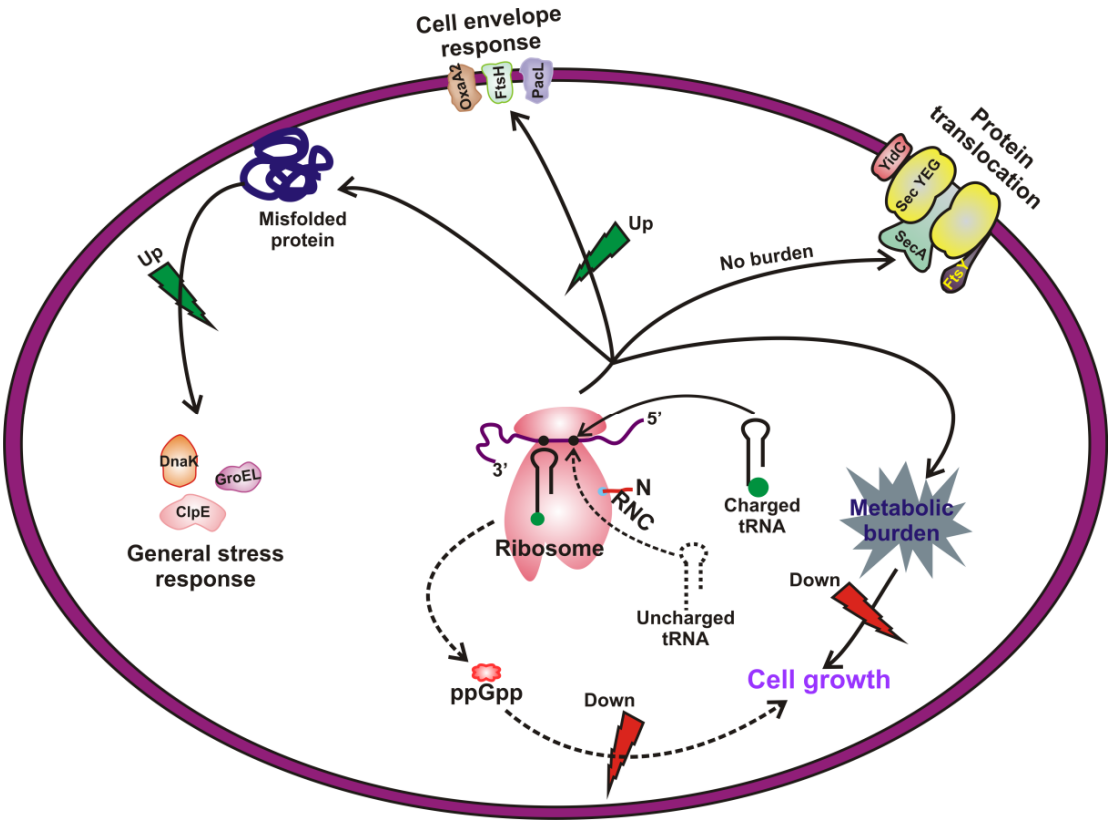


Figure 10. Schematic representation of the major consequences of expressing recombinant (membrane) proteins in *L. lactis*.

Intriguingly, we note that irrespective of the expression levels and/or activity of the overproduced protein the stress response was similar in all cases. The observations on the general and cell envelope stress are also in agreement with a recent study in which the human ABC chloride channel CFTR was expressed in *L. lactis* [59]. In preliminary experiments in which we monitored the effect of overexpression OpuAC, the soluble substrate-binding domain of OpuA, as a cytoplasmic protein, we did not observe the general and cell envelope stress response, that is, neither at the level of the transcriptome nor proteome. This indicates that there is a specific response in *L. lactis* towards membrane protein production, of which the CesSR-mediated one seems most relevant for producing well folded protein (see accompanying paper). In contrast to *E. coli*, overproduction of membrane/secretory proteins in *L. lactis* does not elicit a major change in the expression of the Sec protein translocation machinery On the basis of comparisons with published studies on *L. lactis* physiology and stress response, we conclude many that differences in the expression of genes in for instance carbon, energy

and nucleotide metabolism are due to the decreased growth rate. Under these conditions, the cell may divert nutrients towards the synthesis of recombinant proteins and/or have too limited a capacity to import amino acids in case of nitrogen metabolism. Finally, it is evident that upon membrane protein production the cell is sensing various changes that elicit a response in the transcription machinery, which is not reflected in the proteome. The proteome may thus be a better indicator for cell engineering in order to overcome expression bottlenecks.

Material and methods

Bacterial strains and growth conditions

Lactococcus lactis NZ9000 [60] derivatives were grown at 30°C in M17 broth (Difco, Detroit, MI, USA) containing 1% glucose (w/v) and 5 µg/ml chloramphenicol (GM17-Cm5). *L. lactis* Opu401 is a derivative of *L. lactis* NZ9000 in which the chromosomal *opuA* genes have been deleted by double cross-over recombination [61].

Plasmid construction

DNA manipulations were done according to standard procedures. Plasmids and primer sequences used in this study are listed in Table 2 and 3, respectively.

Table 2. Bacterial strains and plasmids used in this study

Strains	***Description***	***Source***
L. lactis MG1363	*L. lactis* subsp. *cremoris*, plasmid-free derivative of NCDO712	[62]
L. lactis NZ9000	*L. lactis* MG1363Δ*pepN*::*nisRK*	[63]
L. lactis Opu401	*L. lactis* NZ9000Δ*opuAABC*	[64]
Plasmids		
pNZ8048	Cm^r; Expression vector with nisin A-inducible promoter P_{nisA}	[65]
pNZOpuA-His	pNZ8048 containing *opuAA* and *opuABC*	[66]
pNZOpuA(H223A)-His	pNZOpuA derivative; specifying OpuA with the His at position 223 replaced by Ala	This work
pNZ*opuA*(mRNA)-His	pNZOpuA derivative carrying *opuA* with ATG codons at positions 1, 1224 and 1344 replaced by TAA; the numbering refers to the position of nucleotide in the gene sequence	This work
pNZPS1Δ9-His	pNZcLIC derivative containing *PS1Δ9*	This work
pNZStSUT1-His	pNZcLIC derivative containing *StSUT1*	This work
pNZ8902-AmyQ-His	pNZ8902 carrying the gene for α-amylase from *Bacillus amyloliquefaciens* (the protein has a hexa-histidine tag on the C-terminus	Hein Trip, unpublished
pNZAmyQ-His	pNZ8048 containing *amyQ*	This work

The *opuA* mutants, either specifying an inactive version of OpuA or a non-translatable transcript, were constructed by site-directed mutagenesis on pNZOpuA [67]. Ligation-Independent Cloning of *PS1Δ9* and *StSUT1* in pREcLIC and the subsequent conversion of the plasmids into lactococcal expression vectors by Vector Backbone Exchange (VBEx) were performed as described [68]. pNZAmyQ was produced by digesting pNZ8902-*amyQ*

with *Sac*I and *Xba*I and subsequent ligation of the *amyQ-H6* containing fragment into pNZ8048, equally digested with *Sac*I and *Xba*I. All proteins were produced with a C-terminal his-tag for rapid screening of expression. Preparation of electrocompetent cells and electrotransformation into *L. lactis* was performed as described [69].

Table 3. Oligonucleotides used in this study

Name	**Nucleotide sequence 5'→ 3'**
For Site-Directed Mutagenesis	
H223A-FWD	CTTTGTTTCTGCTGATTTGAACGAAG
H223A_REV	ATAATTGTTTTTTGAAATTTTGCTTG
AAstop_FWD	TAAGCAGTAAAAATAAAAATTGAACATTTAAC
AA stop_REV	GGTGAGTGCCTCCTTATAATTTA
ABCstop_FWD	TAAATTGATTTAGCTATTGGACAAGTACC
ABCstop_REV	TTATTCTCCTCCTCTTTTTCAATCTC
ABC(1344)stop_FWD	GGTATTACAGGTGCTTTAACTGCTG
ABC(1344)stop_REV	GTTTTACAATACCGTTCCTGATTTTTG
For RT-qPCR	
opuAA_FWD	TGGTGGGATGCAACAACGGG
opuAA_REV	CGAATCAAAGGGTCAAGAGC
opuABC_FWD	ATGGTTTCTGCTTGGTTACC
opuABC_REV	ATAACTTGGAACAACAAACC
cesSR_FWD	GGATAGACTTGACGGTATTGG
cesSR_REV	TTTTGCTCCTGCAGCCAGAGC
dnaK_FWD	ACCAACTGCCGCAGCTCTTGC
dnaK_REV	ACACCGTCGCCAAGTTCAAGG
bcaP_FWD	GAATGGTCCTTGCAGGTTCG
bcaP_REV	ATGGCTAAGGTCCACACACC
PS1Δ9_FWD	TGGTGGTCGTGGCTACC
PS1Δ9_REV	AGTGCAGGGCTCTCTGG
stsut1_FWD	GGCTCTTCAGCTCTCTTTGC
stsut1_REV	CTGTAGTAGCCGACAACTGG

Protein production and immunodetection

For protein (over)production, cells were grown until OD_{600} ≈ 0.5 and induced with 10^{-3} volume of filtered culture supernatant of the nisin-A producing strain *L. lactis* NZ9700, containing 10 µg/L of nisin A (hereafter referred to as nisin solution). Cells were allowed to grow for the required amount of time and harvested by centrifugation. Sample volumes were normalized on the basis of OD so that an equivalent amount of whole-cell protein was taken for all samples. Cells were washed once with 100 mM potassium phosphate (KPi), pH 7.0, and resuspended in 1 ml of ice-cold 100 mM KPi, pH 7.0, 10% glycerol (w/v), 1 mM $MgSO_4$, 1 mM PMSF and trace amounts of DNAse I. After the addition of 300 mg of glass beads (~100 µm diameter), cells were lysed by three rounds of bead beating in a Fastprep machine for 20 seconds (speed 6.0) with cooling intervals of 5 min on ice in between. Unbroken cells and cell debris were removed by centrifugation at 16,100 x *g* for 30 min and membrane fragments were collected by centrifugation at 267,000 x *g* for 20 min. Protein samples were resolved on 12.5% SDS-PAGE gels and detected by immunodetection with an anti-histidine tag primary monoclonal antibody (GE Healthcare, Uppsala, Sweden). Chemiluminescence detection

was done using the Western-light kit with CSPD (Tropix Inc, Bedford, MA, USA) as the substrate and imaging with the LAS-3000 imaging system (Fujifilm, Minatoku, Tokyo, Japan).

Reverse Transcriptase – Quantitative PCR

An equivalent of 10 OD_{600} units (OD_{600}*V (ml)) of *L. lactis* cells were harvested by centrifugation and cell pellets were kept at -80ºC until further processing. Cells were washed with DEPC treated $T_{10}E_1$ buffer and resuspended in 500 µl $T_{10}E_1$ (10 mM Tris-HCl pH 8.0, 1 mM Na_2-EDTA) and transferred to 2 ml screw-cap tubes. To this cell suspension, 50 µl 10% SDS (w/v), 500 µl phenol/chloroform, 500 mg glass beads (50-105 µm of diameter) and 175 µl macaloid suspension (Bentone MA, Hightstown, NJ) were added. All reagents used for RNA work were treated with diethylpyrocarbonate, (DEPC) Sigma-Aldrich, St. Louis, MO). The macaloid suspension was prepared as follows: 2 g Macaloid was added to 100 ml $T_{10}E_1$, boiled for 5 min, then cooled to room temperature and sonicated for short periods of time until a gel was formed; the gel was spun down in a microcentrifuge and resuspended in 50 ml $T_{10}E_1$ (pH 8.0). Cells were disrupted by bead-beating twice for 45 sec in a Mini-BeadBeater (Biospec Products, Bartlesville, OK) with a 1-min cooling interval on ice. The cell lysate was cleared by centrifugation and 500 µl supernatant was extracted with 500 µl phenol/chloroform, and subsequently with 500 µl chloroform. Total RNA was isolated from the water phase using the High Pure RNA Isolation Kit (Roche Molecular Biochemicals, Mannheim, Germany), according to the manufacturer's protocol. RNA quality was verified with an Agilent Bioanalyzer 2100 using RNA 6000 LabChips (Agilent Technologies Netherlands BV, Amstelveen, the Netherlands) and RNA concentration was determined spectrophotometrically with a Nanodrop ND1000 (NanoDrop Technologies, Wilmington, DE). Copy DNA (cDNA) was synthesized using Superscript III Reverse Transcriptase (Invitrogen, Carlsbad, CA) and the quantification was done with Maxima SYBR Green qPCR Master Mix (Fermentas, Burlington, Canada), according to the suppliers instructurions and using an optical iCyler (BioRad, Hercules, California, USA). The forward and reverse primers used for the qPCR were spaced 100 bp apart, around the 5' regions of the transcripts (Table 3). All reactions were done with identical amounts of RNA to allow comparison of different time points. The data was analyzed as described previously [70].

Glycine betaine transport assay

Whole-cell glycine betaine transport was measured essentially as described before [71]. Cells were grown in GM17 to an OD_{600} ≈ 0.5 and induced for 1 h with 100 ng/L nisin A (final concentration), harvested (at OD_{600} ≈ 1.0) and washed with 50 mM K-HEPES pH 7.3, concentrated to OD_{600} ≈ 50 and kept on ice until use. For the transport assay, the cells were diluted to OD_{600} of 5.0 into 50 mM K-HEPES pH 7.3 plus 650 mM sucrose and 10 mM glucose and pre-energized for 5 min by incubation at 30˚C; the sucrose imposes hyperosmotic conditions which activates the transporer [72]. The assay was started by the addition of ^{14}C-glycine betaine to a final concentration of 15 µM; the transport

reaction was stopped at given times by the addition of 2 ml of ice cold stop buffer (650 mM sucrose in 50 mM K-HEPES pH 7.3), followed by filtration through 450 nm pore size nitrocellulose filters. The filters were washed twice with 2 ml of stop buffer and subsequently transferred into vials containing 2 ml of scintillation fluid. The radio activity was measured using a TriCarb-2800 TR liquid scintillation analyzer (PerkinElmer, Massachusetts, USA).

Preparation of cell samples for transcriptomics and proteomics

To assure true biological replicates, *L. lactis* cells transformed with appropriate plasmids were streaked onto GM17-Cm5 agar plates and single colonies were used to start pre-inoculums in GM17-Cm5 medium. After 8-10 h of growth of the pre-inoculums, several dilutions were prepared, ranging from 10^{-2} to 10^{-6} fold, to obtain an overnight culture growing exponentially (*i.e.*, OD_{600} = 0.2-0.4). This culture was used to inoculate 2.5 L of fresh GM17-Cm5 medium (1/100 dilution). The culture was stirred at 200 rpm and the pH was maintained at 6.8 by automatic addition of 4 M KOH. Cells were grown until $OD_{600} \approx 0.5$ and then induced by the addition of 10 µg/L of nisin A.

For mRNA isolation, an equivalent of 10 OD_{600} units (OD_{600}*V (ml)) of *L. lactis* cells was harvested at the various time points. Cells were quenched in -40ºC cold methanol (the relative final volume of the sample being 60%) to inhibit further mRNA synthesis and degradation. After all samples had been obtained and kept at -40ºC, cells were centrifuged for 3 min at 12,000 x g and a temperature of 4ºC. The pellets were resuspended in ice-cold 500 µl $T_{10}E_1$, transferred to 2 ml screw-cap tubes, immediately frozen in liquid nitrogen and kept at -80°C until further processing.

For proteome analysis, one liter of culture sample was rapidly withdrawn, supplemented with chloramphenicol (100 µg/ml, final concentration) to inhibit protein synthesis. Cells were harvested by centrifugation at 8281 x g for 15 min at 4°C. The cell pellet was washed once with ice-cold 100 mM Kpi, pH 7.0, resuspended in 5 ml of the same buffer, frozen in liquid nitrogen and stored at -80°C.

Proteome analysis

Cells were lysed by three passes through a pre-cooled small French Press cell (Thermo IEC, Waltham, MA, USA) at 12,500 psi. Whole cells and cell debris were removed by centrifugation at 7,650 x *g* for 15 min and membrane fragments were collected by subsequent centrifugation at 267,000 x *g* for 30 min. The supernatant containing the soluble proteome was aliquoted to 500 µl, frozen in liquid nitrogen and stored at -80°C. The pellet fraction containing the membrane proteome was resuspended in 100 mM KPi, pH 7.0, plus 20% glycerol (w/v) and aliquoted to 500 µl, frozen in liquid nitrogen and stored at -80°C. Prior to proteome analyses, the protein concentration of the samples was determined using the 2D-quant kit (GE Healthcare, Uppsala, Sweden).

2D gel electrophoresis and protein identification

The soluble proteome analysis was analyzed by two-dimensional differential in gel electrophoresis (2D DIGE) as described before [73]. For identification, protein spots of interest (average intensity ratio greater than 1.5 and a t-value <0.01) were picked into a 96 well plate using the Ettan spot picker (GE Healthcare, Uppsala, Sweden) equipped with a 2 mm diameter picker head. Gel plugs were washed with 50% acetonitrile (v/v) in 50 mM ammonium bicarbonate, followed by dehydration with 100% acetonitrile (v/v). The gel plugs were first allowed to dry and were subsequently re-swollen in 2 µl of a 15 ng/µl trypsin solution (Trypsin Gold, Promega, Madison, WI, USA) prepared in 40 mM ammonium bicarbonate, 10 % acetonitrile (v/v). After one hour incubation at 37°C, 5 µl of 40 mM ammonium bicarbonate was added and the gel plugs were left overnight at 37°C. For peptide extraction, 20 µl of 1% trifluoroacetic acid (TFA) (v/v) were added and the samples were sonicated in a waterbath for 10 min. The extracts were transferred to a clean plate, and the peptides were further extracted from the gel plugs with 30% acetonitrile (v/v) in 1 % TFA (v/v) and again with 70% acetonitrile (v/v) in 0.1 % TFA (v/v). The extracted peptides were vacuum dried and subsequently resuspended in 5% formic acid (v/v). Peptides were analyzed by nanoLC-MS/MS using an LTQ-Orbitrap XL (Thermo Fisher) electrospray ionization-linear ion trap mass spectrometer. Peptide samples were desalted on a pre-column (C18 PepMap 100, 5 µm, 100Å, 300 µm I.D.×5 mm, LC Packings) and separated on a capillary reverse-phase column (Reverse phase column: C18 PepMap 100, 3 µm, 100 Å, I.D. 75 µm ×150 mm, LC Packings) connected to a Proxeon Easy-LC system. Peptides were eluted by applying a gradient solvent of 5 to 50% (Solvent system: A, 0.1% formic acid (v/v); solvent B, 100% acetonitrile/0.1% formic acid (v/v)) during 30 min and directly analyzed in the LTQ-XL. MS spectra were collected in a range from 400 to 1700 m/z, and the five most abundant ions were submitted to fragmentation (35% normalized collision energy). The peptide mass profiling data were submitted to Mascot (Matrix Science) and searched against the *L. lactis* MG1363 database [74]. The database was created by combining forward and reversed entries of the *L.lactis MG1363* database (release version 31.08.07) plus extra sequences of porcine trypsin (P00761), human keratins (P35908, P35527, P13645, NP_006112), the plasmid related proteins *repA, repC* and chlorampenicol transferase (Q48138, Q48693, P00485), and the overproduced proteins OpuA from *L.lactis IL1403* (Q9KIF7 (OpuAA)/Q7DAU8 (OpuABC)), AmyQ from *B. amyloliquefaciens* (P00692), StSut1 from *Solanum tuberosum* (Q43653) and the human PS1Δ9 (P49768), yielding a database with a total of 4905 entries. Peptide tolerance was set to 10 ppm and 0.8 Da for intact peptides and fragment ions, respectively, and allowing for 1 missed trypsin cleavage. Oxidation of methionine residues and deamidation of asparagine and glutamine were specified as variable modifications. The MS/MS based peptide and protein identifications were further validated with the program Scaffold (version Scaffold_2_04_00, Proteome Software Inc., Portland, OR). Protein identifications established at greater than 99.0%

probability, based on at least 2 peptides identified independently by MS/MS with probability higher than 95%, were accepted.

Membrane protein extraction

To remove the majority of soluble protein contaminants from the membranes, the membrane vesicles were centrifuged at 272,000 x *g* for 20 min at 4°C, and the membranes were resuspended in 50 mM KPi pH 7.0 (=buffer A) to a final concentration of 10 mg/ml (based on protein concentration determination using Bradford reagent, with bovine serum albumin as calibration standard) in a volume of 200-500 µl (initial volume) and kept on ice. An equal volume of buffer A supplemented with 10 M urea plus 10 mM K-EDTA was added slowly, while being stirred and the solution was incubated on ice for 20 min. The 'stripped' membrane vesicles were collected by centrifugation at 272,000 *x g* for 1 h at 4°C and resuspended in the initial volume with buffer A. An equal volume of buffer A supplemented with 12% cholic acid (w/v) was added slowly to the membrane vesicles, while being stirred and the solution was incubated for 20 min on ice. The membrane vesicles were collected by centrifugation at 272,000 *x g* for 1 h at 4°C and resuspended in the initial volume with buffer A. Protein concentrations were on average around 1 mg/ml after the two extraction steps.

Trypsin digestion and iTRAQ labeling of the membrane proteome

For the iTRAQ labeling of the peptides, 50 µg (8-plex) or 100 µg (4-plex) of the extracted membrane vesicles was freeze-dried and resuspended in 20 µl dissolution buffer (500 mM triethyl ammonium-bicarbonate pH 8.5) plus 0.1% SDS (w/v). Reduction of disulfide bonds with tris-(2-carboxyethyl)phosphine (TCEP), cysteine blocking with methyl methanethiosulfonate (MMTS dissolved in isopropanol), digestion with trypsin (Promega) and iTRAQ labeling were done as described by the manufacturers protocol (Applied Biosystems) with minor changes: trypsin digested samples were centrifuged at 16,100 x g for 10 min at room temperature, and the supernatant was collected and stored for iTRAQ labeling. The pellet was resuspended in 10 µl dissolution buffer plus 0.1% SDS (w/v) and digested with trypsin (as described above) for 4 h at 37 °C to achieve maximal cleavage of all proteins. The solution was centrifuged at 16,100 x g for 10 min at room temperature, and the supernatant fractions of both trypsin digestions were combined and dried under vacuum to a volume of approximately 20 µl. The iTRAQ labels were resuspended in 200 µl isopropanol (8-plex) or ethanol (4-plex) instead of the 50 µl isopropanol (8-plex) or 70 µl ethanol (4-plex) as described in the labeling protocol to ensure a high percentage of organic solvent during labeling. The iTRAQ labels of the 8-plex were splitted over two different samples (thus, 100 µL was used per reaction). 4-plex iTRAQ labels were used for the analysis of the membrane proteome of OpuA(H223A) (label 115 & 117) and the *opuA*(mRNA) controls (label 114 & 116), using membrane vesicles from two different cultures (biological replicates). 8-plex iTRAQ labels were used for the analysis of the membrane proteome of the control 8048 (label

113 & 114) AmyQ (label 115 & 116), StSUT1 (label 117 & 118) and PS1Δ9 (119 & 121), using the membrane vesicles of two different cultures (biological replicates).

Separation of iTRAQ-labeled peptides by strong cation exchange (SCX) and reverse phase nano-liquid chromatography (RP-nLC)

iTRAQ labeling was checked, prior to combining the samples. Approximately 200 µg of the combined iTRAQ labeled peptides was dried under vacuum to reduce the volume and then resuspended to a total volume of 600 µl in 10 mM triethyl ammonium phosphate (TEAP) pH 2.7, plus 25% acetonitrile (v/v) (=buffer S1). The pH of the sample was set to pH 2.7 with 1 M phosphoric acid prior to loading of 500 µL of the iTRAQ labeled peptide mixture on the strong cation exchange column [PolyLC Inc. Columbia USA, column volume (CV) =0.7 ml], mounted on an ÅKTA purification system (Amersham Biosciences). The gradient over the column was set up using buffer S1 and buffer S1 plus 500 mM KCl (=buffer S2). The column was prewashed with 5 CV buffer S1, prior to injection of the sample to the column. The column was washed with 10 CV buffer S1 and eluted subsequently with a gradient of 0-5% buffer S2 (in 5 CV, 45 sec fractions collected), a gradient of 5-24% buffer S2 (in 5 CV, 45 sec fractions collected), a gradient of 24-60% (in 5 CV, 1 minute fractions collected) and 5 CV of 100% buffer S2 and 3 CV of 100% buffer S1. Fractions were collected in 96-well plates and the fractions were dried (or reduced to a small volume) in the speedvac and resuspended in a total volume of 50 µl 0.1% trifluoric acid (v/v). Fractions at the end of the gradient were pooled, yielding around 60 different fractions for separation over a reverse-phase nano-liquid chromatography (RP-nLC) column. For the separation of peptides on RP-nLC, peptides were trapped on a pre-column (300 µm x 5 mm, 5 µm particles PepMap C18, p/n 160454) and subsequently separated over an analytical C18 column (75 µm x 150 mm, 3 µm particles 100 Å pores, Acclaim PepMap100 C18) using the Ultimate 3000 nano LC system (Dionex). The gradient over the column was set up using 0.05% TFA (v/v) (=buffer RP1) and 80% acetonitrile (v/v) plus 0.05% TFA (v/v) (=buffer RP2), using a flow rate of 300 nl/min. Column equilibration and sample loading was done in 4% buffer RP2, the gradient was run from 4-40% buffer RP2 over 50 min, 40-60% buffer RP2 in 10 min and from 60-100% buffer RP2 in 5 min. Collection of the eluted peptides was started after 15 min by mixing 1:4 with 2.4 mg/ml α-cyano-4-hydroxy-cinnamic acid matrix (LaserBio Labs). Fractions of 12 seconds were spotted on MALDI targets (260 spots per nLC run) using a Probot system (LC Packings, Amsterdam, The Netherlands). Peptides were analyzed with a 4800 plus MALDI-TOF/TOF Analyzer (Applied Biosystems, Foster City, CA – USA). The MALDI-TOF/TOF was operated in reflectron positive-ionization mode in the m/z range 900-4000. MS plates were calibrated, using 8 calibration spots around the sample set with a calibration mixture of 6 peptides in the range m/z of 904-3657; this calibration was used for the default plate calibration. LC runs were split in two portions to reduce the number of measurements: even numbered LC runs were analyzed at a low m/z range and the odd numbered LC runs were analyzed at the high m/z range. For the

even LC runs: MS/MS precursors were selected from the top 15 peaks (S/N 100+) in the range of m/z 900-2000, and for the odd LC runs: MS/MS precursors were selected from the top 10 peaks (S/N 50) in the range of m/z 2000-4000. The MS/MS spectra were acquired using 2 kV acceleration voltage and air as collision gas at $1x10^{-6}$ torr. A precursor mass transition window of 300 Full Width of Half Maximum (FWHM) (m/z 900-2000) or 200 FWHM (m/z 2000-4000) was used. A peak list of the acquired MS/MS spectra were generated using default settings and a threshold level of S/N 8.

Protein identification of the membrane proteome

MS/MS peak lists were extracted using ProteinPilot (version 2.0.1) and proteins were identified with Mascot analysis (in ProteinPilot version 2.0.1), using a database created as described in the previous section for protein identifications from 2D gel plugs. Scaffold software (version: Scaffold_2_04_00) was used to validate MS/MS-based peptide and protein identifications, using the following settings: fragment tolerance 0.4 Da (monoisotopic), parent tolerance 200 ppm (monoisotopic), fixed modifications +46 for Cys (MMTS), +144 for 4-plex iTRAQ and +304 for 8-plex iTRAQ for Lys and N-terminus, variable modifications +1 for Asn and Gln (deamidation), + 16 for Met (oxidation), and + 144 for 4-plex iTRAQ) and +304 for 8-plex iTRAQ for Tyr and trypsin as digestion enzyme. Two search engines were used: Mascot (Matrix Science, London, UK; version Mascot) with maximal 1 cleavage allowed and X!-tandem (version 2007.01.01.1) with maximal 2 cleavages allowed. Peptide identifications were accepted with a probability above 95% (as specified by the Peptide Prophet algorithm [75]) and proteins were accepted when the probability score was higher than 99% (as specified by the Protein Prophet algorithm [76]) and at least two unique peptides were identified. Proteins containing similar peptides that could not be differentiated by MS/MS analysis alone were grouped to satisfy the principles of parsimony. No hits of the reverse database were found in the list of proteins obtained from Scaffold.

Relative quantification of the membrane proteome

Relative quantification of proteins was based on peptides labeled with iTRAQ reagents [Applied Biosystems, 4-plex (mass +144 Da) or 8-plex (mass +304 Da)] at the amino-group on the N-terminus and at the lysine side chains of the peptides. The reporter ions were fragmented in the MS/MSMS mode (4-plex m/z range 114-117, 8-plex m/z range 113-119 and 121) and the peak areas of these reporter ions were exported to ProteinPilot (version 2.0.1) for relative quantification. The software applies automated settings specific for measurements with the MALDI 4800 to define the mass tolerances and to correct for isotopic impurities of the iTRAQ labels. Peptides that matched multiple proteins were excluded from the quantification. A global bias correction was done for each label, which was calculated on the assumption that the sum of all peak areas per label should be 1/4 (4-plex) or 1/8 (8-plex) of the total. All measured peak areas were then corrected with this bias correction factor, weight averaged and relative protein ratios were calculated using the Paragon algorithm [77], together with a

probability of the ratios (p-values ranging from 0 to 1). Differences in relative protein ratios were considered to be significant if both ratios of the biological replicates had a p-value below 0.01 and a $\log_2$-ratio of at least +/- 0.3 and a similar effect (up/down-regulation). The cut-offs were chosen based on experiments done with technical replicates, using membrane vesicles as starting material (data not shown).

DNA-microarray experiments

Total RNA isolation and cDNA synthesis was performed as described above for the RT-qPCR, except that amino allyl-modified dUTP's were used in the nucleotide mix for cDNA synthesis. Indirect Cy-3/Cy-5 labelling of cDNA was performed according to supplier's instructions (Amersham Biosciences, Piscataway, NJ). Hybridisation of Cy-labelled cDNA was performed during 16 h at 45°C in a microarray hybridisation incubator ISO20 (Grant Boekel, Cambridgeshire, UK) in Ambion Slidehyb #1 hybridisation buffer (Ambion Biosystems, Foster City, CA). SuperAmine glass slides (ArrayIt, Sunnyvale, CA) spotted with duplicates of around 2500 ORF amplicons of *L. lactis* subsp. *cremoris* MG1363 [78] were used. Slides were scanned using a GenePix Autoloader 4200AL scanner (Molecular Devices Corporation, Sunnyvale, CA). DNA microarray data from biological replicates were obtained through dye-swaps, to discard possible differences between the Cy-3 and Cy-5 labeling reactions. Slide images were analyzed using ArrayPro 4.5 (Media Cybernetics, Silver Spring, MD) and the data processed and normalized using MicroPrep [79,80]. The expression ratios were calculated from measurements of at least 4 spots. Differential expression tests were performed with the Cyber-T implementation of a variant of the *t*-test [81]. Only values with an associated *p*-value lower than 0.01 were considered significant.

Overall data analysis

For analysis, proteins were grouped in functional categories using either the COG (Clusters of orthologous genes; [82,83] or KEGG (Kyoto encyclopedia of genes and genomes; [84,85] annotation. The NCBI website was used to run protein blast searches against the non-redundant database of *L. lactis subsp. cremoris* MG1363 (taxid: 416870). The Venn diagrams were made using information obtained with Venncy (version 1.0), which was kindly provided by Bas van Breukelen (University Utrecht, The Netherlands). The PCA and the diagrams of the mRNA and protein ratios were redrawn in Excel (Microsoft, Redmond, WA) using Genesis [86].

Acknowledgements

We thank M.K. Doeven and H. Trip for providing plasmids. This research was supported by the Netherlands Proteomics Centre (NPC), the European Membrane Protein Consortium EDICT, and the Netherlands Science Foundation (NWO; Chemical Sciences Top Subsidy to BP; grant number 700-56-302).

References

[1] Hopkins,A.L. & Groom,C.R. (2002) The druggable genome. *Nat. Rev. Drug Discov.*, **1,** 727-730.

[2] Wallin,E. & von,H.G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, **7,** 1029-1038.

[3] Grisshammer,R. & Tate,C.G. (1995) Overexpression of integral membrane proteins for structural studies. *Q. Rev. Biophys.*, **28,** 315-422.

[4] Junge,F., Schneider,B., Reckel,S., Schwarz,D., Dotsch,V., & Bernhard,F. (2008) Large-scale production of functional membrane proteins. *Cell Mol. Life Sci.*, **65,** 1729-1755.

[5] Katzen,F., Peterson,T.C., & Kudlicki,W. (2009) Membrane protein expression: no cells required. *Trends Biotechnol.*, **27,** 455-460.

[6] Mierau,I. & Kleerebezem,M. (2005) 10 years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis. *Appl. Microbiol. Biotechnol.*, **68,** 705-717.

[7] Surade,S., Klein,M., Stolt-Bergner,P.C., Muenke,C., Roy,A., & Michel,H. (2006) Comparative analysis and "expression space" coverage of the production of prokaryotic membrane proteins for structural genomics. *Protein Sci.*, **15,** 2178-2189.

[8] Quick,M. & Javitch,J.A. (2007) Monitoring the function of membrane transport proteins in detergent-solubilized form. *Proc. Natl. Acad. Sci. U. S. A*, **104,** 3603-3608.

[9] Morello,E., Bermudez-Humaran,L.G., Llull,D., Sole,V., Miraglio,N., Langella,P., & Poquet,I. (2008) Lactococcus lactis, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.*, **14,** 48-58.

[10] Mulligan,C., Geertsma,E.R., Severi,E., Kelly,D.J., Poolman,B., & Thomas,G.H. (2009) The substrate-binding protein imposes directionality on an electrochemical sodium gradient-driven TRAP transporter. *Proc. Natl. Acad. Sci. U. S. A*, **106,** 1778-1783.

[11] Geertsma,E.R. & Poolman,B. (2007) High-throughput cloning and expression in recalcitrant bacteria. *Nat. Methods*, **4,** 705-707.

[12] Kunji,E.R., Slotboom,D.J., & Poolman,B. (2003) Lactococcus lactis as host for overproduction of functional membrane proteins. *Biochim. Biophys. Acta*, **1610,** 97-108.

[13] de Ruyter,P.G., Kuipers,O.P., & de Vos,W.M. (1996) Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin. *Appl. Environ. Microbiol.*, **62,** 3662-3667.

[14] Llull,D. & Poquet,I. (2004) New expression system tightly controlled by zinc availability in Lactococcus lactis. *Appl. Environ. Microbiol.*, **70,** 5398-5406.

[15] Kunji,E.R., Slotboom,D.J., & Poolman,B. (2003) Lactococcus lactis as host for overproduction of functional membrane proteins. *Biochim. Biophys. Acta*, **1610,** 97-108.

[16] Monne,M., Chan,K.W., Slotboom,D.J., & Kunji,E.R. (2005) Functional expression of eukaryotic membrane proteins in Lactococcus lactis. *Protein Sci.*, **14,** 3048-3056.

[17] Geertsma,E.R., Groeneveld,M., Slotboom,D.J., & Poolman,B. (2008) Quality control of overexpressed membrane proteins. *Proc. Natl. Acad. Sci. U. S. A*, **105,** 5722-5727.

[18] Marreddy,R.K.R., Geertsma,E.R., Permentier,H.P., Pinto,J.P.C., Kok,J., & Poolman,B. (2010) Amino Acid Accumulation Limits the Overexpression of Proteins in Lactococcus lactis. *PLoS ONE*, **5,** e10317.

[19] Locher,K.P., Lee,A.T., & Rees,D.C. (2002) The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science*, **296,** 1091-1098.

[20] Lewinson,O., Lee,A.T., & Rees,D.C. (2008) The funnel approach to the precrystallization production of membrane proteins. *J. Mol. Biol.*, **377,** 62-73.

[21] Miroux,B. & Walker,J.E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.*, **260,** 289-298.

[22] Massey-Gendel,E., Zhao,A., Boulting,G., Kim,H.Y., Balamotis,M.A., Seligman,L.M., Nakamoto,R.K., & Bowie,J.U. (2009) Genetic selection system for improving recombinant membrane protein expression in E. coli. *Protein Sci.*, **18,** 372-383.

[23] Serrano-Vega,M.J., Magnani,F., Shibata,Y., & Tate,C.G. (2008) Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U. S. A*, **105,** 877-882.

[24] Bonander,N., Hedfalk,K., Larsson,C., Mostad,P., Chang,C., Gustafsson,L., & Bill,R.M. (2005) Design of improved membrane protein production experiments: quantitation of the host response. *Protein Sci.*, **14,** 1729-1740.

[25] Wagner,S., Baars,L., Ytterberg,A.J., Klussmeier,A., Wagner,C.S., Nord,O., Nygren,P.A., van Wijk,K.J., & de Gier,J.W. (2007) Consequences of membrane protein overexpression in Escherichia coli. *Mol. Cell Proteomics.*, **6,** 1527-1550.

[26] Wagner,S., Klepsch,M.M., Schlegel,S., Appel,A., Draheim,R., Tarry,M., Hogbom,M., van Wijk,K.J., Slotboom,D.J., Persson,J.O., & de Gier,J.W. (2008) Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A*, **105,** 14371-14376.

[27] Bonander,N., Darby,R.A., Grgic,L., Bora,N., Wen,J., Brogna,S., Poyner,D.R., O'Neill,M.A., & Bill,R.M. (2009) Altering the ribosomal subunit ratio in yeast maximizes recombinant protein yield. *Microb. Cell Fact.*, **8,** 10.

[28] Kuipers,O.P., de Ruyter,P.G.G.A., Kleerebezem,M., & de Vos,W.M. (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology*, **64,** 15-21.

[29] Biemans-Oldehinkel,E. & Poolman,B. (2003) On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. *EMBO J.*, **22,** 5983-5993.

[30] Zaitseva,J., Jenewein,S., Jumpertz,T., Holland,I.B., & Schmitt,L. (2005) H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J.*, **24,** 1901-1910.

[31] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. U. S. A*, **103,** 10624-10629.

[32] Brouwer,R.W.W., Pinto,J.P.C., Zeyniyev,A., Kuipers,O.P., & Kok,J. The growth-phase dependent transcriptome of *Lactococcus lactis.* 2010.

Ref Type: Generic

[33] Kuipers,O.P., de,J.A., Baerends,R.J., van Hijum,S.A., Zomer,A.L., Karsens,H.A., den Hengst,C.D., Kramer,N.E., Buist,G., & Kok,J. (2002) Transcriptome analysis and related databases of Lactococcus lactis. *Antonie Van Leeuwenhoek*, **82,** 113-122.

[34] Newman,M.J., Foster,D.L., Wilson,T.H., & Kaback,H.R. (1981) Purification and reconstitution of functional lactose carrier from Escherichia coli. *J. Biol. Chem.*, **256,** 11804-11808.

[35] Uchida,E., Ohsumi,Y., & Anraku,Y. (1985) Purification and properties of H+-translocating, Mg2+-adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae. *J. Biol. Chem.*, **260,** 1090-1095.

[36] Steen,A., Wiederhold,E., Gandhi,T., Breitling,R., & Slotboom,D.J. (2010) Physiological adaptation of the bacterium Lactococcus lactis in response to the production of human CFTR. *Mol. Cell Proteomics.*.

[37] Gitton,C., Meyrand,M., Wang,J., Caron,C., Trubuil,A., Guillot,A., & Mistou,M.Y. (2005) Proteomic signature of Lactococcus lactis NCDO763 cultivated in milk. *Appl. Environ. Microbiol.*, **71,** 7152-7163.

[38] Budin-Verneuil,A., Pichereau,V., Auffray,Y., Ehrlich,D.S., & Maguin,E. (2005) Proteomic characterization of the acid tolerance response in Lactococcus lactis MG1363. *Proteomics.*, **5,** 4794-4807.

[39] Dressaire,C., Redon,E., Milhem,H., Besse,P., Loubiere,P., & Cocaign-Bousquet,M. (2008) Growth rate regulated genes and their wide involvement in the Lactococcus lactis stress responses. *BMC. Genomics*, **9,** 343.

[40] Guedon,E., Sperandio,B., Pons,N., Ehrlich,S.D., & Renault,P. (2005) Overall control of nitrogen metabolism in Lactococcus lactis by CodY, and possible models for CodY regulation in Firmicutes. *Microbiology*, **151,** 3895-3909.

[41] Marreddy,R.K.R., Geertsma,E.R., Permentier,H.P., Pinto,J.P.C., Kok,J., & Poolman,B. (2010) Amino Acid Accumulation Limits the Overexpression of Proteins in Lactococcus lactis. *PLoS ONE*, **5,** e10317.

[42] Schlame,M. (2008) Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. *J. Lipid Res.*, **49,** 1607-1620.

[43] Martinez,B., Zomer,A.L., Rodriguez,A., Kok,J., & Kuipers,O.P. (2007) Cell envelope stress induced by the bacteriocin Lcn972 is sensed by the Lactococcal two-component system CesSR. *Mol. Microbiol.*, **64,** 473-486.

[44] Veiga,P., Bulbarela-Sampieri,C., Furlan,S., Maisons,A., Chapot-Chartier,M.P., Erkelenz,M., Mervelet,P., Noirot,P., Frees,D., Kuipers,O.P., Kok,J., Gruss,A., Buist,G., & Kulakauskas,S. (2007) SpxB regulates O-acetylation-dependent resistance of Lactococcus lactis peptidoglycan to hydrolysis. *J. Biol. Chem.*, **282,** 19342-19354.

[45] van,B.E., Dekker,H.L., Froderberg,L., Houben,E.N., Urbanus,M.L., de Koster,C.G., de Gier,J.W., & Luirink,J. (2008) Detection of cross-links between FtsH, YidC, HflK/C suggests a linked role for these proteins in quality control upon insertion of bacterial inner membrane proteins. *FEBS Lett.*, **582,** 1419-1424.

[46] Chang,D.E., Smalley,D.J., & Conway,T. (2002) Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model. *Mol. Microbiol.*, **45,** 289-306.

[47] Eymann,C., Homuth,G., Scharf,C., & Hecker,M. (2002) Bacillus subtilis functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *J. Bacteriol.*, **184,** 2500-2520.

[48] Rallu,F., Gruss,A., & Maguin,E. (1996) Lactococcus lactis and stress. *Antonie Van Leeuwenhoek*, **70,** 243-251.
[49] Budin-Verneuil,A., Pichereau,V., Auffray,Y., Ehrlich,D., & Maguin,E. (2007) Proteome phenotyping of acid stress-resistant mutants of Lactococcus lactis MG1363. *Proteomics.*, **7,** 2038-2046.
[50] van Asseldonk M., Simons,A., Visser,H., de Vos,W.M., & Simons,G. (1993) Cloning, nucleotide sequence, and regulatory analysis of the Lactococcus lactis dnaJ gene. *J. Bacteriol.*, **175,** 1637-1644.
[51] Eaton,T., Shearman,C., & Gasson,M. (1993) Cloning and sequence analysis of the dnaK gene region of Lactococcus lactis subsp. lactis. *J. Gen. Microbiol.*, **139,** 3253-3264.
[52] Varmanen,P., Ingmer,H., & Vogensen,F.K. (2000) ctsR of Lactococcus lactis encodes a negative regulator of clp gene expression. *Microbiology*, **146 (Pt 6),** 1447-1455.
[53] Kilstrup,M., Jacobsen,S., Hammer,K., & Vogensen,F.K. (1997) Induction of heat shock proteins DnaK, GroEL, and GroES by salt stress in Lactococcus lactis. *Appl. Environ. Microbiol.*, **63,** 1826-1837.
[54] Budin-Verneuil,A., Pichereau,V., Auffray,Y., Ehrlich,D.S., & Maguin,E. (2005) Proteomic characterization of the acid tolerance response in Lactococcus lactis MG1363. *Proteomics.*, **5,** 4794-4807.
[55] Martinez,B., Zomer,A.L., Rodriguez,A., Kok,J., & Kuipers,O.P. (2007) Cell envelope stress induced by the bacteriocin Lcn972 is sensed by the Lactococcal two-component system CesSR. *Mol. Microbiol.*, **64,** 473-486.
[56] Martinez,B., Zomer,A.L., Rodriguez,A., Kok,J., & Kuipers,O.P. (2007) Cell envelope stress induced by the bacteriocin Lcn972 is sensed by the Lactococcal two-component system CesSR. *Mol. Microbiol.*, **64,** 473-486.
[57] Veiga,P., Bulbarela-Sampieri,C., Furlan,S., Maisons,A., Chapot-Chartier,M.P., Erkelenz,M., Mervelet,P., Noirot,P., Frees,D., Kuipers,O.P., Kok,J., Gruss,A., Buist,G., & Kulakauskas,S. (2007) SpxB regulates O-acetylation-dependent resistance of Lactococcus lactis peptidoglycan to hydrolysis. *J. Biol. Chem.*, **282,** 19342-19354.
[58] Linares,D.M., Geertsma,E.R., & Poolman,B. Evolved *Lactococcus lactis* strains for enhanced expression of recombinant membrane proteins. 2010.
[59] Steen,A., Wiederhold,E., Gandhi,T., Breitling,R., & Slotboom,D.J. (2010) Physiological adaptation of the bacterium Lactococcus lactis in response to the production of human CFTR. *Mol. Cell Proteomics.*.
[60] Kuipers,O.P., de Ruyter,P.G.G.A., Kleerebezem,M., & de Vos,W.M. (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology*, **64,** 15-21.
[61] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. U. S. A*, **103,** 10624-10629.
[62] Gasson,M.J. (1983) Plasmid complements of Streptococcus lactis NCDO 712 and other lactic streptococci after protoplast-induced curing. *J. Bacteriol.*, **154,** 1-9.
[63] Kuipers,O.P., de Ruyter,P.G.G.A., Kleerebezem,M., & de Vos,W.M. (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. *Journal of Biotechnology*, **64,** 15-21.

[64] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. U. S. A*, **103,** 10624-10629.
[65] Kleerebezem,M., Beerthuyzen,M.M., Vaughan,E.E., de Vos,W.M., & Kuipers,O.P. (1997) Controlled gene expression systems for lactic acid bacteria: transferable nisin-inducible expression cassettes for Lactococcus, Leuconostoc, and Lactobacillus spp. *Appl. Environ. Microbiol.*, **63,** 4581-4584.
[66] van der Heide T. & Poolman,B. (2000) Osmoregulated ABC-transport system of Lactococcus lactis senses water stress via changes in the physical state of the membrane. *Proc. Natl. Acad. Sci. U. S. A*, **97,** 7102-7106.
[67] van der Heide T. & Poolman,B. (2000) Osmoregulated ABC-transport system of Lactococcus lactis senses water stress via changes in the physical state of the membrane. *Proc. Natl. Acad. Sci. U. S. A*, **97,** 7102-7106.
[68] Geertsma,E.R. & Poolman,B. (2007) High-throughput cloning and expression in recalcitrant bacteria. *Nat. Methods*, **4,** 705-707.
[69] Holo,H. & Nes,I.F. (1989) High-Frequency Transformation, by Electroporation, of Lactococcus lactis subsp. cremoris Grown with Glycine in Osmotically Stabilized Media. *Appl. Environ. Microbiol.*, **55,** 3119-3123.
[70] Nolan,T., Hands,R.E., & Bustin,S.A. (2006) Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.*, **1,** 1559-1582.
[71] Mahmood,N.A., Biemans-Oldehinkel,E., & Poolman,B. (2009) Engineering of ion sensing by the cystathionine beta-synthase module of the ABC transporter OpuA. *J. Biol. Chem.*, **284,** 14368-14376.
[72] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. U. S. A*, **103,** 10624-10629.
[73] Marreddy,R.K.R., Geertsma,E.R., Permentier,H.P., Pinto,J.P.C., Kok,J., & Poolman,B. (2010) Amino Acid Accumulation Limits the Overexpression of Proteins in Lactococcus lactis. *PLoS ONE*, **5,** e10317.
[74] Wegmann,U., O'Connell-Motherway,M., Zomer,A., Buist,G., Shearman,C., Canchaya,C., Ventura,M., Goesmann,A., Gasson,M.J., Kuipers,O.P., van,S.D., & Kok,J. (2007) Complete genome sequence of the prototype lactic acid bacterium Lactococcus lactis subsp. cremoris MG1363. *J. Bacteriol.*, **189,** 3256-3270.
[75] Keller,A., Nesvizhskii,A.I., Kolker,E., & Aebersold,R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.*, **74,** 5383-5392.
[76] Nesvizhskii,A.I., Keller,A., Kolker,E., & Aebersold,R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.*, **75,** 4646-4658.
[77] Shilov,I.V., Seymour,S.L., Patel,A.A., Loboda,A., Tang,W.H., Keating,S.P., Hunter,C.L., Nuwaysir,L.M., & Schaeffer,D.A. (2007) The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell Proteomics.*, **6,** 1638-1655.
[78] Kuipers,O.P., de,J.A., Baerends,R.J., van Hijum,S.A., Zomer,A.L., Karsens,H.A., den Hengst,C.D., Kramer,N.E., Buist,G., & Kok,J. (2002) Transcriptome analysis and related databases of Lactococcus lactis. *Antonie Van Leeuwenhoek*, **82,** 113-122.
[79] van Hijum,S.A., de,J.A., Buist,G., Kok,J., & Kuipers,O.P. (2003) UniFrag and GenomePrimer: selection of primers for genome-wide production of unique amplicons. *Bioinformatics.*, **19,** 1580-1582.

[80] van Hijum,S.A., de,J.A., Baerends,R.J., Karsens,H.A., Kramer,N.E., Larsen,R., den Hengst,C.D., Albers,C.J., Kok,J., & Kuipers,O.P. (2005) A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC. Genomics*, **6,** 77.

[81] Long,A.D., Mangalam,H.J., Chan,B.Y., Tolleri,L., Hatfield,G.W., & Baldi,P. (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. *J. Biol. Chem.*, **276,** 19937-19944.

[82] Tatusov,R.L., Koonin,E.V., & Lipman,D.J. (1997) A genomic perspective on protein families. *Science*, **278,** 631-637.

[83] Tatusov,R.L., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Kiryutin,B., Koonin,E.V., Krylov,D.M., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N., Rao,B.S., Smirnov,S., Sverdlov,A.V., Vasudevan,S., Wolf,Y.I., Yin,J.J., & Natale,D.A. (2003) The COG database: an updated version includes eukaryotes. *BMC. Bioinformatics.*, **4,** 41.

[84] Kanehisa,M., Goto,S., Hattori,M., oki-Kinoshita,K.F., Itoh,M., Kawashima,S., Katayama,T., Araki,M., & Hirakawa,M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34,** D354-D357.

[85] Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T., & Yamanishi,Y. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36,** D480-D484.

[86] Sturn,A., Quackenbush,J., & Trajanoski,Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics.*, **18,** 207-208.

Chapter 6 Summary and perspectives

Profiling of adenine nucleotide-binding proteins with a focus on ABC transporters

Chapter 1 describes the structural fold of adenine nucleotide-binding proteins with different functions, which allows their grouping in terms of mode of nucleotide binding. The second chapter presents a number of crystal structures for these different folds and shows how the nucleotides are bound in the different proteins. In order to enrich for the adenine nucleotide-binding proteins, two analogues: *p*-biotinyl-amidobenzoic acid-ATP (*p*-BABA-ATP) and *p*-biotinyl-amidomethyl benzoic acid-ATP (*p*-BAMBA-ATP) were designed and synthesized. The two analogues were designed to fit in the space around the ribose moiety of ATP observed in the crystal structures of ABC transporters. Preliminary screening of complex protein samples shows that identification of proteins is still limited by high background binding and bottlenecks in the separation and detection of peptides (Chapter 3). Further research will have to focus on the optimization of the separation and detection of the peptides, or by synthesizing ATP analogues with an even higher affinity for ABC proteins than *p*-BABA-ATP and *p*-BAMBA-ATP. An option would be to introduce an azido-group on the adenine ring. The azido-group (described in Chapter 2) can be used to cross-link the ATP analog covalently to the nucleotide-binding domains, which enables harsh washing steps to reduce the aspecific binding on the streptactin column.

The ABC transporter OpuA

The ABC transporter OpuA, used in this proteomics study (Chapter 5) as an example of a membrane protein that is well overproduced in *L.lactis,* has been characterized for its function and regulation [1-4]. Chapter 4 presents a detailed biochemical analysis of the substrate-binding domain (termed OpuAC). We also obtained crystal structures of OpuAC (Chapter 4), but crystallization of the whole protein complex has so far not been successful (still ongoing work within our group). Two Master students (Hugo van Beek and Faizah Fulyani), whom I supervised during my PhD, have optimized purification protocols and tried to achieve crystals of the full protein complex. Also, a mutant of OpuA, *i.e*. OpuAE189Q, was constructed, which is predicted to lock the protein in the transition state. A similar mutant in the maltose transporter stabilized the protein for crystallization [5], but so far it did not yet yield crystals of OpuA. In another Masters project, a homolog of OpuA from *Listeria monocytogenes* (lmo_1421/1422) was overexpressed in *L.lactis*. Crystallization trials have been set up for this protein, a project that is now continued by Dr. Andreja Vujicic-Zagar and has led to diffracting crystals albeit at low resolution.

Because OpuA is a complex protein with two subtrate-binding domains, two nucleotide-binding domains, two-times-two CBS domains (hereafter referred to as CBS module) and two amphipathic helices connected to the TMD, we pursued electron

microscopy (EM) to obtain the overall architecture of the transporter at low resolution. For this analysis, fractions of purified OpuA were negatively stained, using uranyl-acetate, and subsequently characterized by single-particle analysis. Figure 1 shows preliminary EM images. The full-length protein shows specific structural features (Figure 1A), with the largest sphere most likely being the membrane-embedded part surrounded by a detergent shell. The two smaller spheres could either be the nucleotide-binding domains plus associated CBS module or the substrate-binding domains. In order to discriminate between these two possibilities, the effects of substrate binding on the structure were determined with the substrates glycine betaine and MgATP (Mg-ATP plus *ortho*-vanadate). Addition of magnesium-ATP (MgATP) did not yield any change in the overall architecture as assessed by the single-particle EM analysis (Figure 1C). However, upon addition of glycine betaine a subpopulation emerged with only one small sphere (Figure 1B). This could indicate that one of the SBDs is now docking onto the membrane domain, whereas in the absence of glycine betaine both SBDs are more distantly linked. The structural changes were specific for glycine betaine and irrespective of whether or not MgATP was present, indicating that the smaller spheres are the substrate-binding domains rather than the nucleotide-binding domains.

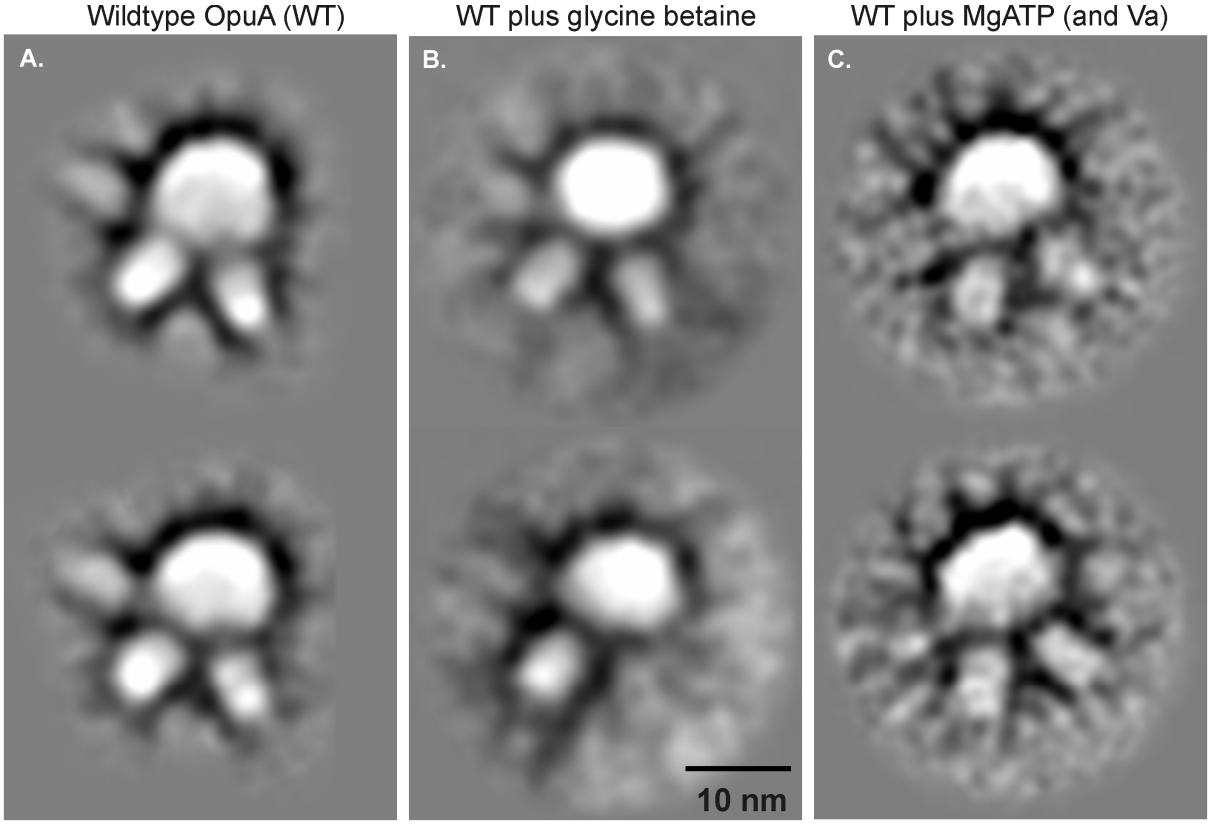


Figure 1. Negatively-stained electron microscopy images of purified OpuA without (panel A), in the presence of glycine betaine (1 mM, panel B) or MgATP (3 mM, panel C). In case of Mg-ATP, *ortho*-vanadate (Va) was also included at a final concentration of 3mM.

Based on these results, a relatively large change in the hydrodynamic properties of the protein is expected upon binding of glycine betaine. The (change in the) shape was

investigated further by ion mobility-mass spectrometry. Detection of intact protein complexes was first measured by nanoflow electrospray mass spectrometry. This technique combines the soft ionization of the electrospray, which can preserve non-covalent interactions, with a nanoflow, which facilitates the desolvation process. For the detection of the full protein complex, purified protein was brought into the gas phase with the detergent micelle still associated with the protein. Subsequent gas phase collisions were tuned to activate and disrupt the detergent micelle but to leave the protein complex and associated ligands intact. In this way, one can obtain accurate information about the mass of the protein complex, the stoichiometry and heterogeneity of the complex, the stability of the protein, and about dynamic interactions between subunits and ligands [6,7]. For purified OpuA, both the subunits, OpuABC and OpuAA (Figure 2A), and the full OpuA complex (Figure 2B) were observed. The measured and calculated molecular weights are listed in Table 1.

Table 1. Calculated and measured masses (with an accuracy of 0.1 kDa) of OpuA and individual subunits, using nanoflow electronspray mass spectrometry.

Protein	**Oligomeric state**	**Measured mass (kDa)**	**Predicted mass (kDa)**	**Mass difference (kDa)**
OpuAA	Monomer	45.8	45.7	0.1
	Dimer	91.4-91.5	91.4	0-0.1
	Tetramer	182.7	182.8	0.1
OpuABC	Monomer	68.5	63.4	5.1
	Dimer	137.1	126.8	10.3
OpuA	Dimer	228.4	218.2	10.2

The analysis of the ATPase subunits (OpuAA) indicates that both monomeric and dimeric species are present. The mass of the transmembrane/substrate-binding domain protein (OpuABC) was calculated to be 63.4 kDa, which is 5.1 kDa lower than measured. Importantly, a difference of two times 5.1 kDa was observed for dimeric OpuABC and the (dimeric) full size OpuA complex. We speculate that the additional mass originates from lipids tightly bound with the hydrophobic protein domains. Given an average mass of 700 Da, approximately 7 lipid molecules would be present per OpuABC subunit. Maldi-MS/MS analysis of the OpuA protein and of the protein after trypsin digestion showed that the additional mass is not due to read-through of the transcript (data not shown). Since the activity of OpuA is highly dependent of non-bilayer lipids and requires anionic lipids for ionic activation, it is possible that lipids are tightly associated with the protein complex [8,9]. If lipids are indeed present, it will be important to determine their molecular nature.

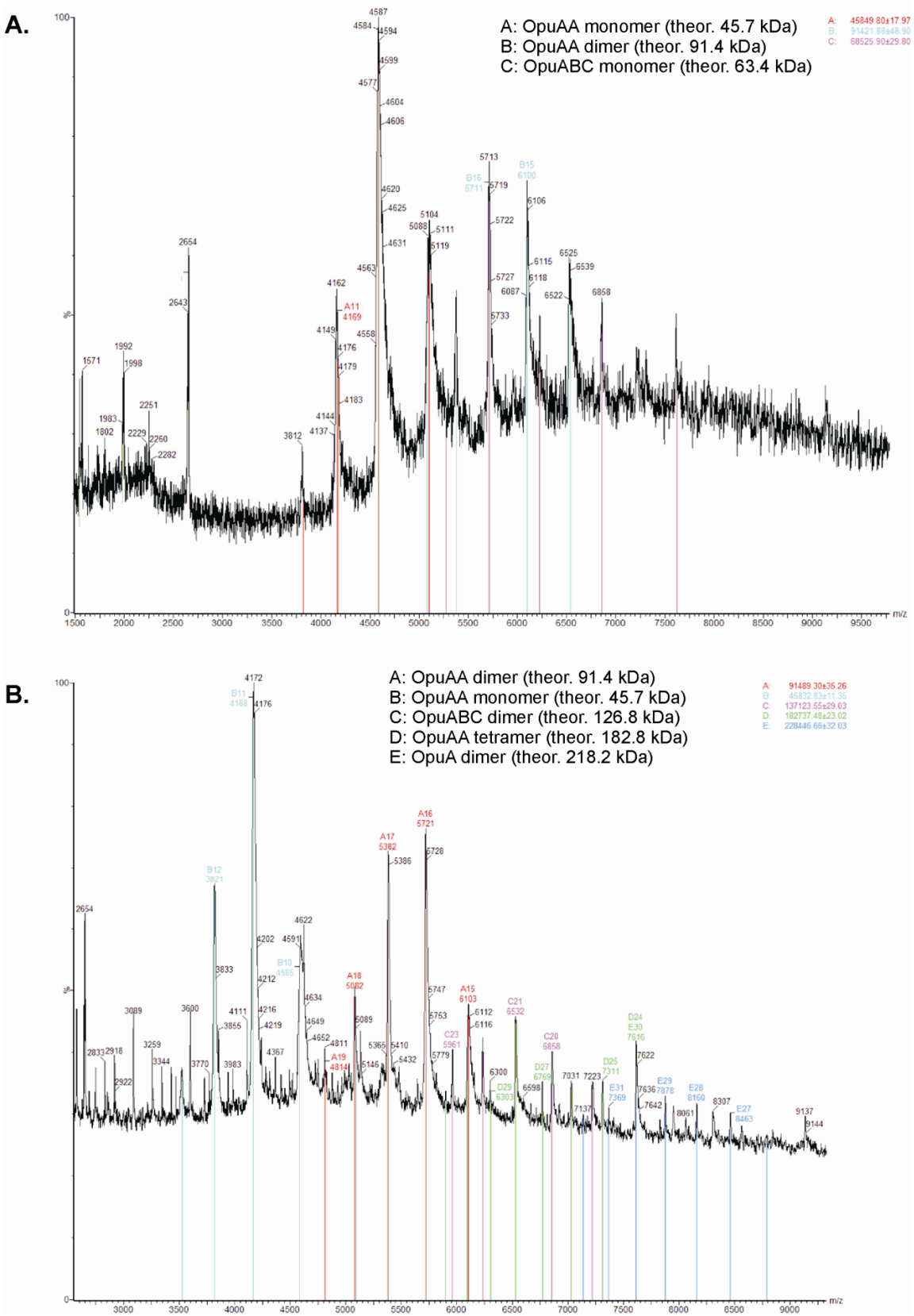


Figure 2. Nanoflow electronspray mass spectrometry (Q-ToF2 MS) analysis of purified OpuA. The OpuA protein was purified as described in the material and methods section and was diluted to a concentration of about 0.5 mg/mL in 200 mM ammonium acetate pH 7.5 (without additional detergent) prior to spraying into the MS. Panel A presents spectra of the individual subunits and Panel B shows spectra of the full size OpuA complex.

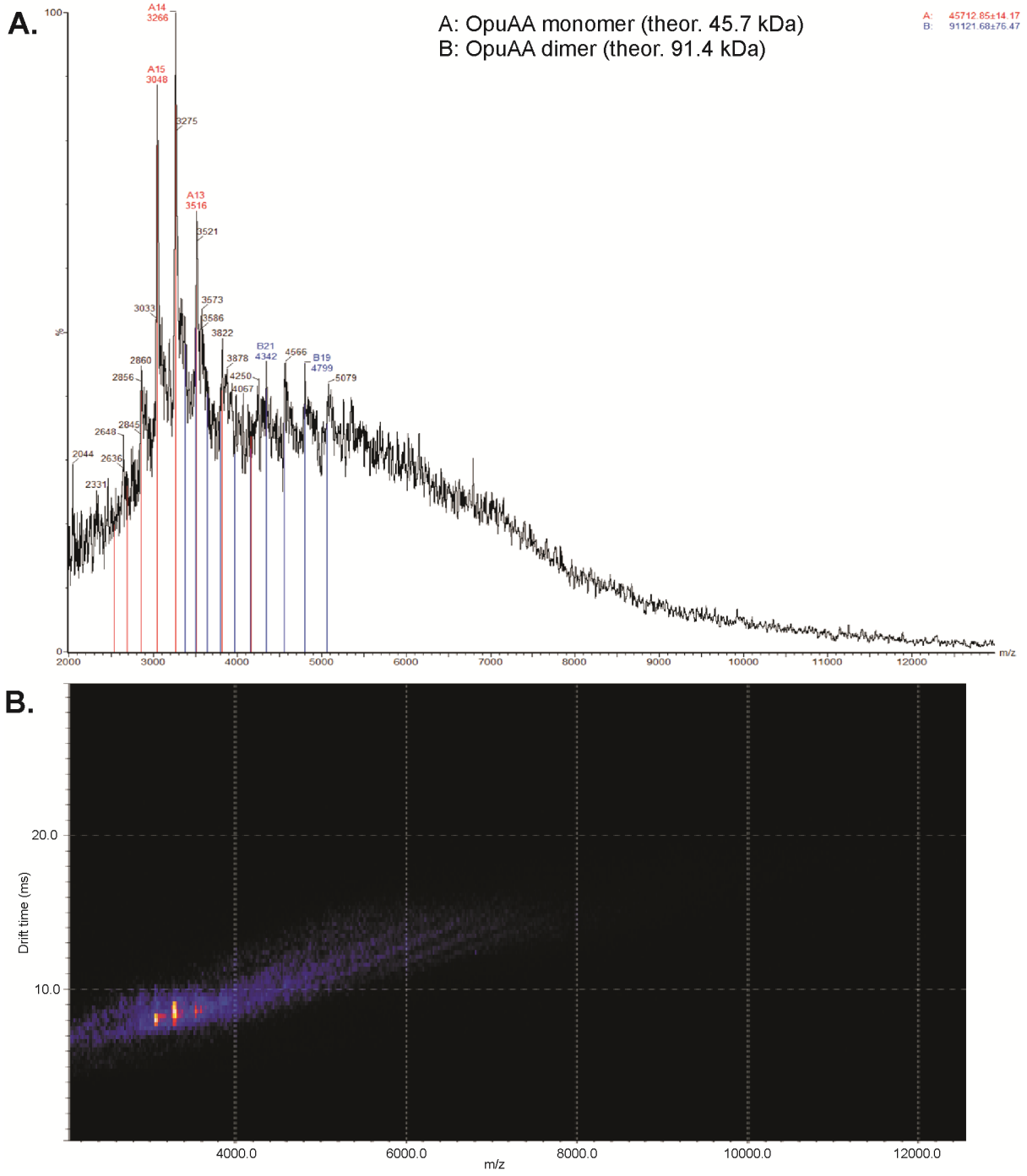


Figure 3. Ion mobility-mass spectrometry analysis of purified OpuA. OpuA was purified as described in the materials and methods section and was diluted to a concentration of ~2 mg/mL in 200 mM ammonium acetate pH 7.5 (without additional detergent) prior to spraying into the MS. The MS spectra (panel A) are similar to the measurements shown in Figure 2. The drift patterns (ion mobility measurements) are measured simultaneously (panel B). The intensity of the m/z-peaks are indicated by the different colors (with yellow being the highest intensity); the drift times (y-axis) are an indication of the ion mobility. Only the drift times of the OpuAA monomer were well defined in these initial measurements.

The results presented in Figure 2 show that the full size OpuA complex can be detected by nanoflow electrospray MS, which is not trivial for complex membrane proteins. The next step will be to perform ion mobility mass spectrometry (IM-MS). In ion mobility measurements, the (macro)molecular ions traverse neutral gas molecules under a weak electrical field. This provides information about the shape and size of the macromolecules, because extended ions will collide more frequently with the gas molecules than compact structures [10]. On the basis of the electron microscopy images significant changes in ion mobility (drift times) are expected upon glycine betaine binding to OpuA. Initial measurements show that it is possible to obtain a good IM-MS signal with purified OpuA (Figure 3), which is already an accomplishment for membrane proteins. The drift time of the OpuAA subunit is clearly detected (Figure 3B), but conditions need to be optimized to resolve the full size OpuA complex.

Material and Methods

Constructs, growth and protein purification

Membrane vesicles containing OpuA [11] or OpuAΔSBD [12] were prepared from *L. lactis* OpuA401 overexpressing the corresponding proteins, which were purified on Ni-Sepharose as described previously [8]. This purification protocol is based on the use of 50 mM KP_i pH 7.0, 200 mM KCl, 20% glycerol plus 0.05% DDM as buffer and addition of imidazole to to elute the proteins from the Ni-Sepharose resin.

Electron microscopy single particle analysis

For electron microscopy single particle analysis, the 50 mM KP_i pH 7 in the purification buffers [8] was replaced by 20 mM K-MES pH 7. Ni-Sepharose-purified OpuA was chromatographed further on a Superdex 200 10/300 Gl column in 20 mM K-MES pH 7.0, 200 mM KCl plus 0.04 % DDM. The glycerol was omitted from the buffer, because it interfered with the electron microscopy analysis. Fractions containing OpuA were pooled and concentrated to 5-10 µM, using a Vivaspin column with a cut-off of 100 kDa. 1 mM glycine betaine or 3 mM Mg-ATP plus 3 mM *ortho*-vanadate were added when appropriate. For electron microscopy analysis, the OpuA protein was diluted in water prior to immobilization of the sample on glow-discharged carbon-coated copper grips. The grids were stained with 2% (w/v) uranyl acetate. The electron microscope (Philips CM120) was operated at 120 kV and images were recorded with a Gatan 4K slow-scan CCD camera at 100.000 magnification with a pixel size of 3.75Å at the specimen level. The "GRACE" software was used for semi-automated specimen selection and data acquisition. Approximately 25.000 single-particle projections were selected per protein sample and single-particle analysis was done with the Groningen Image Processing (GRIP) software package on a personal computer cluster. Selected single-particle projections (96x96 pixel frame) were aligned by multi-reference alignment and reference-free alignment procedures as described previously [13,14]. The particles were subjected to multivariate statistical analysis, followed by hierarchical ascendant

classification [14]. The best projection from each dataset was averaged and summed, after several cycles of multi-reference alignments, statistical analysis and classification.

Mass spectrometry analysis

For the mass spectrometry analysis, the 50 mM KP_i pH 7 plus 200 mM KCl of the purification buffer was replaced by 250 mM ammonium acetate pH 7.5. Fractions containing purified protein were concentrated on the Vivaspin column and subsequently diluted two times with 200 mM ammonium acetate pH 7.5 to reduce the concentration of detergent micelles, glycerol and imidazole. Mass spectrometry (MS) measurements were carried out on a nanoflow Q-ToF2 (Waters, Manchester, UK) instrument. Nanoflow electrospray capillaries were prepared as described previously [15]. The following MS parameters were typically used: capillary voltage, 1.5 kV; sample cone, 150-200 V; collision voltage, 150-200 V. The pressure in the quadrupole was 4.5 X 10-5 mbar. All mass spectra were processed with Masslynx 4.1 software (Waters, UK).

Preliminary ion-mobility mass spectrometry (IM-MS) measurements were carried out on a Synapt HD mass spectrometer (Waters). Typical instrument parameters were: needle voltage 1.5 kV, MCP 2350 V, cone voltage 150-200V and Trap voltage 100-200 V.

Acknowledgements

The electron microscopy images were obtained in collaboration with Kasia Moscicka and Marc Stuart, from the Electron Microscopy group, headed by Prof. E.J. Boekema (University Groningen, NL). The mass spectra were obtained in collaboration with Sheila Wang from the Department of Chemistry headed by Prof. C.V. Robinson (Oxford University, UK). I would like to thank this group for welcoming me to work in their laboratory for several weeks.

References

[1] Biemans-Oldehinkel,E., Mahmood,N.A., & Poolman,B. (2006) A sensor for intracellular ionic strength. *Proc. Natl. Acad. Sci. USA*, **103,** 10624-10629.

[2] Mahmood,N.A., Biemans-Oldehinkel,E., Patzlaff,J.S., Schuurman-Wolters,G.K., & Poolman,B. (2006) Ion specificity and ionic strength dependence of the osmoregulatory ABC transporter OpuA. *J. Biol. Chem*, **281,** 29830-29839.

[3] Patzlaff,J.S., van der Heide,T., & Poolman,B. (2003) The ATP/Substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem*, **278,** 29546-29551.

[4] van der Heide,T. & Poolman,B. (2000) Glycine betaine transport in Lactococcus lactis is osmotically regulated at the level of expression and translocation activity. *J. Bacteriol*, **182,** 203-206.

[5] Oldham,M.L., Khare,D., Quiocho,F.A., Davidson,A.L., & Chen,J. (2007) Crystal structure of a catalytic intermediate of the maltose transporter. *Nature*, **450,** 515-521.

[6] Barrera,N.P., Isaacson,S.C., Zhou,M., Bavro,V.N., Welch,A., Schaedler,T.A., Seeger,M.A., Miguel,R.N., Korkhov,V.M., van Veen,H.W., Venter,H.,

Walmsley,A.R., Tate,C.G., & Robinson,C.V. (2009) Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat. Methods*, **6,** 585-587.

[7] Sharon,M. & Robinson,C.V. (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev Biochem*, **76,** 167-193.

[8] Wolters,J.C., Berntsson,R.P., Gul,N., Karasawa,A., Thunnissen,A.M., Slotboom,D.J., & Poolman,B. (2010) Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLoS. One*, **5,** e10361.

[9] van der,H.T., Stuart,M.C., & Poolman,B. (2001) On the osmotic signal and osmosensing mechanism of an ABC transport system for glycine betaine. *EMBO J*, **20,** 7022-7032.

[10] Ruotolo,B.T. & Robinson,C.V. (2006) Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem Biol*, **10,** 402-408.

[11] van der Heide,T. & Poolman,B. (2000) Osmoregulated ABC-transport system of Lactococcus lactis senses water stress via changes in the physical state of the membrane. *Proc. Natl. Acad. Sci. USA*, **97,** 7102-7106.

[12] Biemans-Oldehinkel,E. & Poolman,B. (2003) On the role of the two extracytoplasmic substrate-binding domains in the ABC transporter OpuA. *EMBO J.*, **22,** 5983-5993.

[13] Penczek,P., Radermacher,M., & Frank,J. (1992) Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, **40,** 33-53.

[14] van Heel,M., Gowen,B., Matadeen,R., Orlova,E.V., Finn,R., Pape,T., Cohen,D., Stark,H., Schmidt,R., Schatz,M., & Patwardhan,A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev Biophys*, **33,** 307-369.

[15] Hernandez,H. & Robinson,C.V. (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nat. Protoc*, **2,** 715-726.

Nederlandse samenvatting voor niet-biologen

Bacteriën en eiwitten

Het onderzoek dat in dit proefschrift gepresenteerd wordt gaat over eiwitten die in een bacterie zitten. Een bacterie wordt gevormd door één cel en is zo klein dat het niet met het blote oog te zien is, maar het heeft tal van belangrijke functies in het milieu en ons lichaam. De bacterie die in dit proefschrift gebruikt is heet *Lactococcus lactis*, een bacterie die leeft op planten. Deze bacterie wordt in de industrie gebruikt voor het maken van o.a. karnemelk en kaas.

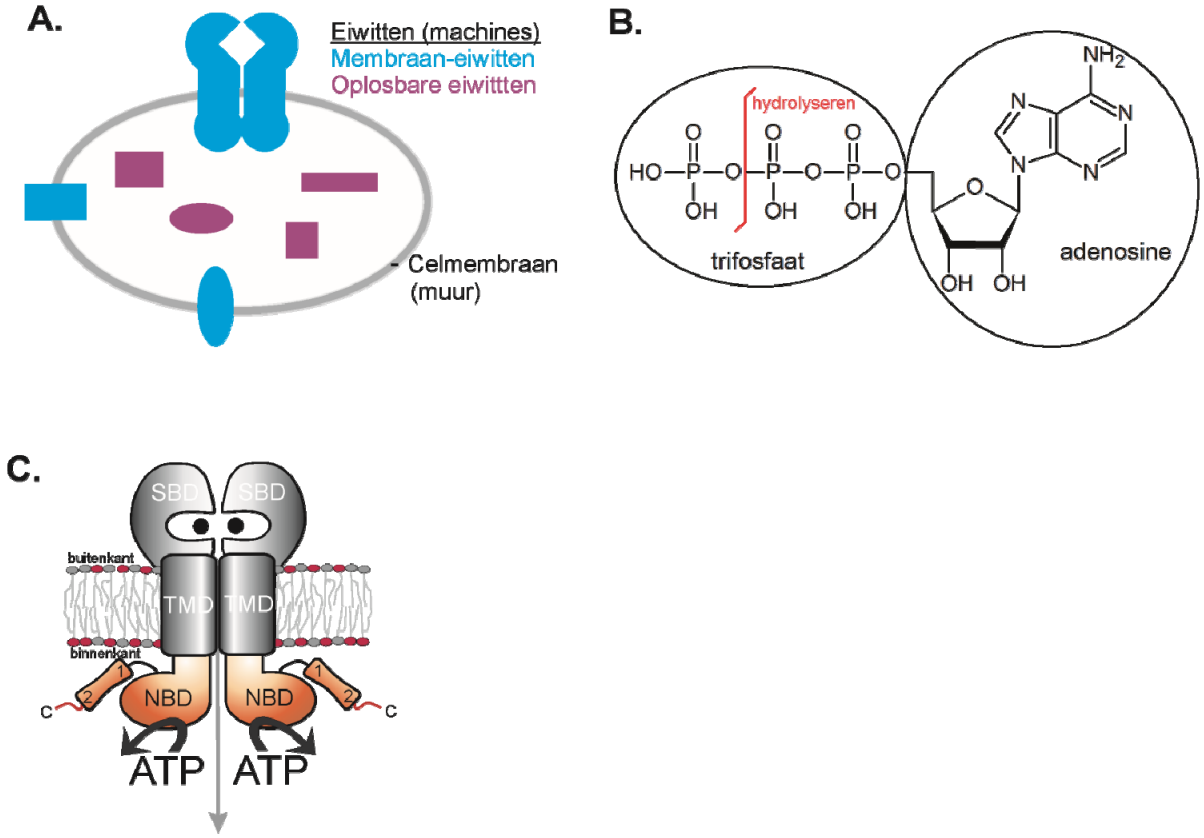
Een bacteriële cel zou je kunnen vergelijken met een fabriek. In de fabriek worden grondstoffen in meerdere stappen omgezet naar een product. Een verschil met een gewone fabriek is dat de cel voornamelijk producten maakt die het nodig heeft voor eigen processen en groei, waaronder het maken van enzymen en motoreiwitten (de machines van de cel). Om het (productie)proces draaiende te houden en de kwaliteit van de producten te garanderen zijn er ook machines die een uitgebreid controlesysteem vormen. Voor de productie zijn meerdere machines nodig met ieder hun eigen functie binnen het proces; er is informatie nodig welke producten er met welke machines gemaakt moeten worden; de kwaliteit van de producten moet in elke fase gecontroleerd en eventueel gecorrigeerd worden en er is energie nodig om de machines te laten draaien. De fabriek wordt door een muur afgescheiden van de buitenwereld, die in bacteriën bestaat uit een celmembraan en celwand (‘bacteriële muur’).

De celwand geeft de stevigheid en bescherming aan een bacteriële cel en de celmembraan zorgt ervoor dat de vloeistof die binnen in de cel aanwezig is met alle machines en andere stoffen gescheiden blijft van de omgeving en vormt de belangrijkste permeabiliteitsbarière. De celmembraan houdt de machines in de cel, maar kleine ongeladen deeltjes zoals water kunnen hier wel doorheen.

De officiële naam voor de machines in een bacteriële cel is enzymen, eiwitten met een katalytische functie: de eiwitten die in de cel in de vloeistof drijven heten oplosbare eiwitten. De eiwitten drijven niet alleen in de vloeistof, een deel van de machines bevindt zich in de celmembraan en deze worden membraaneiwitten genoemd (Figuur 1A). De membraaneiwitten kunnen bijvoorbeeld stoffen de cel in en uit transporteren, ze vormen als het ware de deuren. Een algemeen overzicht van een bacteriële cel is te zien in Figuur 1A. Zo zijn er veel soorten eiwitten die elk een eigen functie hebben, bijvoorbeeld bij het omzetten van grondstoffen in verschillende eindproducten, maar eiwitten zijn ook betrokken bij het transport van de stoffen en communicatie en regulatie van verschillende processen.

Adenosine trifosfaat (ATP)

De machines/eiwitten hebben energie nodig om te kunnen functioneren. De belangrijkste energiebron in de cel is een stof genaamd adenosine trifosfaat (ATP). Energie is opgeslagen in de bindingen van ATP. De structuur van ATP is zien in Figuur 1B. De rechterkant van het ATP molecuul met de twee vijfhoeken en een zeshoek wordt adenosine genoemd en is belangrijk om voor de herkenning door het eiwit dat energie nodig heeft. De linkerkant van de stof ATP zijn de drie (tri) fosfaat-groepen. Fosfaat groepen bestaan uit een fosfor (wordt aangegeven met de letter P) en vier zuurstof-atomen (aangegeven met een O). De energie is opgeslagen in de binding tussen de fosfaat-groepen. Het verbreken van de binding tussen de fosfaatgroepen wordt hydrolyse genoemd en de vrijgekomen energie kan vervolgens gebruikt worden door het eiwit dat de ATP gebonden heeft.



Figuur 1. Paneel A laat een voorbeeld zien van een bacteriële cel. De cel wordt omgeven door een membraan (grijze lijn). De membraaneiwitten zijn in blauw aangegeven en de eiwitten in het cytoplasma in paars. Paneel B laat de structuur van adenosine trifosfaat (ATP) zien. De plek waar de binding van ATP meestal wordt gehydrolyseerd is aangegeven met een rode lijn. Paneel C laat de opbouw van het ABC transporteiwit OpuA zien. Aan de buitenkant zitten de substraat-bindende domeinen (SBD). Het substraat (aangegeven als zwarte bolletjes) wordt vervolgens via de transmembraan domeinen (TMD) de cel in getransporteerd (het volgt de grijze pijl). De energie komt vrij door de hydrolyse (het verbreken van een fosfaat-binding, aangegeven met een rode lijn in paneel B) van ATP door de nucleotide bindende domeinen (NBD), die aan de binnenkant van de cel zitten; twee domeinen die de activiteit van OpuA reguleren staan aangegeven met 1 en 2.

ABC transporteiwitten

De energie van ATP kan gebruikt wordt voor productie van stoffen, controle van processen en voor transport-doeleinden; een overzicht van de verschillende ATP-verbruikende eiwitten staat beschreven in hoofdstuk 1, waaronder de groep ATP-bindende cassette eiwitten (hierna aangeduid als ABC transporters). Deze categorie eiwitten gebruikt de energie uit ATP om stoffen over de celmembraan te transporteren. De algemene opbouw van deze eiwitten wordt uitgelegd in Figuur 1C. Alle ABC transporters hydrolyseren ATP op dezelfde manier om stoffen (veelal substraten genoemd) naar binnen (nutriënten) dan wel naar buiten (eindproducten, toxische verbindingen) te transporteren. Wat er aan nutriënten naar binnen wordt getransporteerd hangt af van de domeinen aan de buitenkant van de cel: de substraat-bindende domeinen. ABC transporters zijn in de regel selectief voor een specifiek substraat of enkele nauw verwante stoffen. Tussen de ATP-bindende domeinen aan de binnenkant van de cel en de substraat-bindende domeinen aan de buitenzijde zitten de transmembraan domeinen: deze domeinen zorgen voor een pad in de celmembraan om het substraat te transporteren. Dat deze transporters belangrijk zijn kun je zien aan de ziektes die veroorzaakt worden als een van de ABC transporters niet meer of niet goed werkt, ziektes als taaislijmziekte of resistentie tegen antibiotica.

Het onderzoek

Omdat de ABC transporters zo belangrijk zijn is het cruciaal om te begrijpen hoe de eiwitten er precies uit zien en hoe ze werken. Het proefschrift beschrijft drie verschillende invalshoeken om deze categorie eiwitten te bestuderen.

De eerste invalshoek staat in hoofdstuk 2 en 3 en beschrijft de ontwikkeling van een methode om ABC transporters uit een bacteriële cel te halen en te scheiden van alle andere componenten. Hoofdstuk 2 beschrijft de ontwikkeling van een molecuul dat de ABC transporters bindt. Het molecuul is zo ontwikkeld dat het niet alleen een gedeelte bevat om ABC transporters te herkennen (namelijk ATP), maar ook een gedeelte waarmee het molecuul uit de oplossing gevist kan worden. Dit principe wordt geïllustreerd in Figuur 1 uit hoofdstuk 3. Van deze op ATP-gebaseerde moleculen zijn twee verschillende versies gemaakt. In hoofdstuk 3 zijn de twee versies van de moleculen getest op hun vermogen om ABC transporters specifiek uit een complexe eiwitmengsels te halen.

De tweede invalshoek richt zich specifiek op de ABC transporter OpuA, waarvan een beschrijving is te vinden in box 1. Hoofdstuk 4 beschrijft hoe goed of slecht het substraat-bindingsdomein van OpuA verschillende substraten kan binden, hoe de transporter geactiveerd wordt en hoe het substraat-bindende domein er op atomair niveau uitziet. Om te meten wat er met het substraat gebeurt, is er gebruik gemaakt van een radio-actief gemerkte variant. Via de straling die het radio-actieve substraat afgeeft kun je namelijk meten waar het substraat blijft: als het substraat aan het eiwit bindt

blijft het eiwit radio-actieve straling afgeven nadat je het ongebonden substraat hebt verwijderd, terwijl het eiwit geen straling afgeeft als het substraat niet gebonden is. Het onderzoeken van de gedetailleerde structuur van OpuA is niet gemakkelijk, omdat de eiwitten niet te zien zijn met een gewone microscoop. Een techniek genaamd X-ray kristallografie kan helpen. In deze techniek gebruik je het feit dat een lichtbundel wordt verstrooid als er iets in de lichtbundel staat. Zo krijg je dus een beeld van het voorwerp dat je niet direct kunt zien onder de microscoop. Om eiwitten met deze techniek te bestuderen moeten ze eerst kristallen (in een vaste stof met een herhalend patroon) vormen en dat is niet vanzelfsprekend vooral voor membraaneiwitten. In hoofdstuk 4 is het gebruik van deze techniek beschreven voor het substraat-bindende domein van OpuA. Voor het hele eiwit is ook geprobeerd om kristallen te krijgen, maar dit is nog niet gelukt. In het perspectief van dit proefschrift (hoofdstuk 6) worden daarom alternatieve methodes beschreven om meer inzicht te krijgen van de structuur van het hele eiwit.

Box 1 De ABC transporter OpuA

In dit proefschrift staat onderzoek beschreven naar een ABC transporteiwit, genaamd OpuA, van de bacterie *Lactococcus lactis*; vergelijkbare eiwitten komen in nagenoeg alle bekende microorganismen voor Deze transporteiwitten worden geactiveerd om de organismen te beschermen tegen hoge zoutconcentraties in het milieu. Als de zoutconcentratie in de omgeving hoger wordt dan zal water aan de cel worden onttrokken en zal deze krimpen. Bij het krimpen gebeuren er twee dingen: de cel kan uitdrogen en door het verlies aan water wordt de druk op de celmembraan lager. De druk op de celmembraan heet de turgordruk en is onder andere belangrijk om de cel gezond te houden (denk maar aan een ballon, als daar de lucht uitgaat verliest het al zijn stevigheid). Daarom wordt bij té grote verschillen de ABC transporter OpuA geactiveerd om de cel te beschermen. OpuA importeert glycine betaine, wat tot molaire concentraties in de cel opgenomen kan worden. Hierdoor neemt dewaterconcentratie in de cel weer toe en wordt de turgordruk op niveau gehouden.

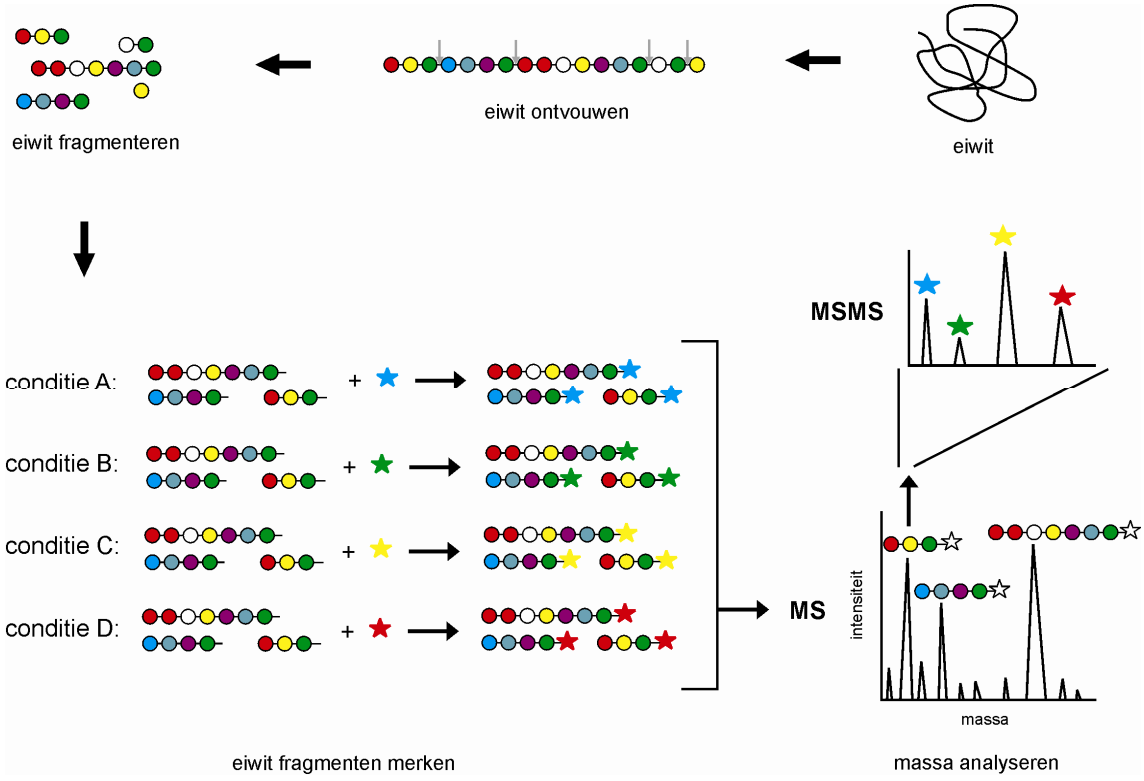
OpuA heeft de algemene opbouw van ABC importers (zoals zichtbaar is in Figuur 1C): aan de buitenkant van de cel zitten de twee substraat-bindende (receptor) domeinen waar de stof die getransporteerd moet worden eerst bindt (de te transporteren stof heet een substraat). In de celmembraan zitten twee transmembrane domeinen die een ruimte creëeren, waardoor glycine betaine specifiek naar binnen kan worden getransporteerd. Aan de binnenkant van de cel zitten de twee ATP-bindende domeinen, die verantwoordelijk zijn voor de hydrolyse van ATP en daamee de transportreactie drijven. De selectiviteit van OpuA is te vergelijken met een slot dat alleen met de juiste sleutel geopend kan worden. Het juiste substraat (glycine betaine) opent het slot, zodat de transportreactie kan plaatsvinden.

Om de ABC transporter OpuA te bestuderen moet het eiwit eerst gemaakt worden en dat wordt gedaan door de bacterie *Lactococcus lactis* opdracht te geven het eiwit in grote hoeveelheden te synthetiseren. *Lactococcus lactis* verwerkt deze opdracht goed voor OpuA, maar dit werkt niet altijd met andere eiwitten. Hoofdstuk 5 beschrijft wat mogelijke knelpunten zijn voor de synthese van membraaneiwittten in *Lactococcus lactis.* Hiervoor is gebruik gemaakt van een techniek genaamd massa spectrometrie, welke beschreven staat in box 2.

Samengevat beschrijft dit proefschrift de ontwikkeling van een methode om meerdere ABC eiwitten in parallel te isoleren en karakteriseren, alsmede de identificatie van bottlenecks bij de synthese van membraaneiwitten.

Box 2 Massa spectrometrie

Eiwitten zou je kunnen zien als (gevouwen) kralenkettingen. De kralen heten aminozuren en in de natuur worden slechts 20 verschillende aminozuren gebruikt voor de synthese van eiwitten. De combinatie en volgorde van de aminozuren en de wijze van vouwing maakt elk eiwit uniek. Dit unieke karakter van eiwitten wordt gebruikt in massa spectrometrie (afgekort: MS): eiwitten worden in stukjes geknipt en de massa (feitelijk de massa gedeeld door de lading) van elk fragment wordt gemeten. Een computer vergelijkt de gemeten massa's met die van een database op basis van de berekende massa van fragmenten van alle eiwitten in een bepaald organisme. Zo kan dus bepaald worden welke eiwitten aanwezig zijn in een cel, maar de techniek kan verder uitgebreid worden door verschillende condities met elkaar te vergelijken. Nadat de eiwitten voor elke conditie in fragmenten geknipt zijn, worden alle fragmenten gekoppeld aan een klein merkje. Voor elke conditie wordt een verschillend merkje gebruikt (in Figuur 2 aangegeven met sterretjes). Deze merkjes zijn zo ontwikkeld dat ze allemaal dezelfde massa hebben wanneer ze in de massa spectrometer worden geanalyseerd. Het bijzondere van de merkjes is dat ze in de massa spectrometer ook gebroken kunnen worden, waarbij elk fragment een unieke massa krijgt en dus een tweede dimensie in de massa spectrometrische analyse oplevert; deze methode wordt MSMS genoemd). Door eiwitten van verschillende cellen (mutanten, groei-condities) te labelen met verschillende merkjes, en te analyseren met behulp van MSMS, kan een beeld verkregen worden van de relatieve concentraties ervan.



Figuur 2. Massa spectrometrische analyse van eiwitten. Eiwitten worden geïdentificeerd via de exacte massa van eiwit-fragmenten (gemeten in de MS analyse). De relatieve hoeveelheden van deze eiwitten kunnen tegelijkertijd vergeleken worden voor verschillende condities, door de eiwit-fragmenten van elke conditie te labelen met een verschillend merkje (verschillende kleuren sterretjes), die daarna gemeten worden met MSMS.

List of publications

Wolters JC, Poolman B (2010) Synthesis and characterization of nucleotide analogs, designed for profiling of ATP-binding proteins (manuscript in preparation).

Marreddy, RKR, Pinto J*, Wolters JC**, Kok J, Poolman B (2010) Response of *Lactococcus lactis,* to membrane protein over production stress (manuscript being submitted).

Wolters JC, Berntsson RP**, Gul N, Karasawa A, Thunnissen AM, Slotboom DJ, Poolman, B (2010). Ligand binding and crystal structures of the substrate-binding domain of the ABC transporter OpuA. *PLos One* **5**(4): e10361.

Folgering JH, Wolters JC, Poolman B, (2005). Engineering covalent oligomers of the mechanosensitive channel of large conductance from *Escherichia coli* with native conductance of gating characteristics*. Protein Sci* **14** (12): 2947-54.

Wolters JC, Abele R, Tampé R, (2005). Selective and ATP-dependent translocation of peptides by the homodimeric ATP binding cassette transporter TAP-Like (ABCB9). *J Biol Chem* **280** (25): 23631-6.

Nawoord

Bijna vijf jaar bezig geweest, maar hier is het dan toch: een afgerond proefschrift! Bert, hartelijk dank voor de kans om m'n promotie-onderzoek in jouw groep te doen. Dank voor de ruimte die je me gaf om m'n onderzoek zelf op te bouwen, het te verbreden en mezelf ook op veel vlakken te kunnen ontwikkelen. Ook een speciaal woord van dank voor jou, Dirk: je maakte altijd tijd om te luisteren en advies te geven, zelfs al was ik niet een van je eigen AIO's.
Ook dank voor de begeleiders van mijn master-projecten: Joost, met jouw enthousiasme is het onmogelijk om er niet door aangestoken te worden. The drive from Rupert and Robert (in Frankfurt) to bring the scientific research to the next level worked highly motivating as well.
Thanks to all the colleagues in our lab, and especially the people I collaborated with in a number of projects: with Ronnie, Nadia and Akira on the OpuA-projects; with Elena, Ravi and Joao on the proteomics/transcriptomics project; with my master-students Faizah, Hugo and Lidiya; with Kasia and Marc on the EM analysis, and with Sheila on the MS analysis in Oxford.
Thanks to the people from the nicest office on Poolman-north (especially with the Xmas lights)! Maarten en Siva, we have shared all the ups and downs of the research and writing at the same time and I enjoyed being in one office with you guys. It's fun to see that we are even defending our theses in the same month. Gea, bedankt voor je luisterend oor, voor de vele gesprekken en natuurlijk voor alle gezelligheid ook na werktijd! De vele gezamelijke etentjes en vooral de vele spelletjes kolonisten met veel verschillende collega's (in het bijzonder de vaste kern: Gea, Inga, Tejas en Astri). Hopelijk gaan we daar nog lang mee door!

Familie en vrienden, het uitleggen waar ik precies mee bezig was is misschien niet altijd gelukt (wie weet helpt de nederlandse samenvatting nog een beetje), maar ik ben blij dat er desondanks zoveel mensen om me heen staan. De steun en gezelligheid van mijn familie is heel belangrijk voor mij. Lekker genieten van de vakanties, weekendjes weg en het vele wandelen met o.a. Marianne, Berend, Arjan & Carolien, Sandra, 2x Jan en Irene & Bert. Simone, reuze gezellig dat je ook in Groningen ging studeren en er bent blijven hangen (zelfs al is dat maar tijdelijk: je overleeft mijn 'kookkunsten' nu alweer 10 jaar, dus het mag wel duidelijk zijn dat onze vriendschap tegen een stootje kan). Henk en Tineke, dank voor jullie vriendschap.

Het laatste woord is de tekst uit het wapen van de universiteit: *VER/BUM/DNI LU/CER/NA:* oftewel *VERBUM DOMINI LUCERNA PEDIBUS NOSTRIS* (Psalm 119:105).